March 2021

Predictive Modeling of FMOL Health System Utilization Using Machine Learning Algorithms and Retrospective Study of COVID Tested Patients

RamyaKrishna Tummala
Louisiana State University and Agricultural and Mechanical College

PREDICTIVE MODELING OF FMOL HEALTH SYSTEM UTILIZATION USING MACHINE LEARNING ALGORITHMS & RETROSPECTIVE STUDY OF COVID TESTED PATIENTS

A Thesis

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Master of Engineering Science

in

Department of Engineering Science

by
RamyaKrishna Tummala
B.S.M.E, VRSEC, 2013
May 2021

Acknowledgements

I would like to thank Our Lady of the Lake Regional Medical Center (OLOL RMC) for providing me an opportunity to be part of this research.

I would like to express my deepest gratitude to my advisor Dr. Gerald Knapp for guiding me in the right direction to complete my master's research successfully. His support, motivation and great knowledge helped me a lot in the entire research work and to enhance my skills as a data analyst.

I would like to thank Dr. Craig Harvey, Dr. Laura Ikuma and, Dr. Tonya Jagneaux for being my committee members and providing their valuable suggestions.

I would like to express my deepest appreciation to Dr. Tonya Jagneaux for providing me the opportunity to use her time and help whenever needed in the research work.

I would like to express my sincere gratitude to Dr. Diana Hamer for her patience, support and technical guidance throughout the research.

Abstract

Overutilization of Emergency Departments (ED) is a major problem among the health care providers in the United States. In this research, a machine learning-based predictive model for predicting ED high utilizers will be designed based on a set of existing and proposed facilities and the population and social determinant of health (SDOH) factors influencing utilization. The purpose of the model will be to alert the healthcare systems and government organizations by identifying the reasons for overutilization of the medical services among the people in a particular community. Also, the novel coronavirus disease 2019 (COVID-19) developed in Whunan city, China has spread quickly to the other parts of the world. It has become a serious health threat to the United States. Moreover, in this research study, the clinical and social characteristics that are responsible for the quick spread of COVID-19 disease across the Louisiana state will be identified. The purpose of this study is to identify what kind of population gets COVID 19 and providing real time care decisions to minimize the risk of an individual acquiring COVID-19. The patient data from Electronic Health Records (EHR) of Francis Missionaries of our Lady Health System (FMOLHS) is geocoded and mapped into ArcGIS software. The socioeconomic factors and social vulnerability Index (SVI) variables available from various online sources are joined to the geocoded patient data with the help of spatial joining techniques available in the ArcGIS software. Correlation analysis between the dependent variables and factors will be conducted.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Abbreviations

ACS	American Community Survey
AdaBoost	Adaptive Boosting
AHRQ	Agency for Healthcare Research and Quality
ANN	Artificial Neural Networks
AUC	Area Under Curve
BR	Baton Rouge
CDC	Centers for Disease Control and Prevention
CHF	Congestive Heart Failure
CHR	Community Health Records
CNI	Community Needs Index
COVID-19	Coronavirus Disease 2019
CPMI	Charlson Predictive Mortality Index
DT	Decision Trees
EBRP	East Baton Rouge Parish
ED	Emergency Departments
EHR	Electronic Health Records
EMR	Electronic Medical Records
FMOLHS	Franciscan Missionaries of Our Lady Health System
FN	False Negative
FP	False Positive
GIS	Geographic Information System

HCUP	Healthcare Cost and Utilization Project
HH	High-High
HIE	Health Information Exchange
HIPAA	Health Insurance Portability and Accountability Act
HL	High-Low
HRSA	Health Resources & Service Administration
ICD	International Classification of Diseases
LA	Louisiana
LH	Low-Low
LL	Low-Low
LOS	Length of Stay
LPG	Lake Physician Group
LR	Logistic Regression
LSU	Louisiana State University
LSUHSC	Louisiana State University Health Science Center
MAE	Mean Absolute Error
MSE	Mean Squared Error
NCHS	National Center for Health Statistics
OLOL RMC	Our Lady of the Lake Regional Medical Center
PCAST	President's Council of Advisors on Science and Technology
PPV	Positive Predictive Value
R^2-Score	R-Squared measure
RMSE	Root Mean Squared Error

RNN	Recurrent Neural network
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SDOH	Social Determinants of Health
SES	Socioeconomic status
SVI	Social Vulnerability Index
TN	True Negative
TP	True Positive
UC	Urgent Care
US	United States
VHR	Veterans Health Administration
WHO	World Health Organization
XGBoost	Extreme Gradient Boosting

Chapter 1. Introduction

1.1. Background

Franciscan Missionaries of Our Lady Health System (FMOLHS) is one of the leading health care in Louisiana and Missippi with 1,747 licensed beds and has ten hospitals within the system [1]. Our Lady of the Lake Regional Medical Center (OLOL RMC) in Baton Rouge has two hospitals associated with it and serves more than 850K patients annually [2]. OLOL clinics are located throughout Louisiana, although there is a significant concentration in the Baton Rouge metropolitan area. The Our Lady of the Lake Physician Group (LPG) provide many services such as emergency services, primary care, cancer care, diabetes & nutrition center, critical care to the people living in Louisiana [3]. OLOL is affiliated with Louisiana State University Health Network (LSUHN), made up of members of Louisiana State University Health Science Center (LSUHSC) across LA. LSUHN provides 30 specialties (Allergy & Immunology, Behavioral Science Center, Cardiology, Neurology, Surgery, and so on) to meet patient healthcare needs. LSUHN consists of 9 clinics across LA with 1 clinic in Baton Rouge [4]. A common Electronic Health Record (EHR) is maintained by all FMOLHS affiliated hospitals and clinics including OLOL RMC.

The primary motivation for this thesis was the development of template procedures for FMOLHS analysts to integrate Geographic Information Systems (GIS) and geospatial analysis and machine learning into their research. In collaboration with OLOL RMC staff, two geospatial-based research problems of current interest to the FMOLHS were identified, each covering different elements of analysis. Each of the problems addressed are also important research contributions in their own right.

For one of the problems, we designed a machine learning-based predictive classification model for predicting ED high utilizers based on the multiple ED facilities available in LA and the population and social determinant of health (SDOH) factors influencing utilization. The purpose of the model is to help healthcare systems and government organizations in identifying the reasons for overutilization of the medical services among the people in a particular community. For decades, Emergency Departments (ED) use in the United States (US) has been increasing steadily at a rate faster than the US population growth [5]. According to the Healthcare Cost and Utilization Project (HCUP), ED high utilizers are defined as patients with 4 or more visits per year to the ED [6].

There is a dramatic increase of 66.4% and 28.5% in the number of ED visits covered by Medicaid and Medicare respectively between 2006 and 2014 [7]. Hospital closures in a particular area or region are one of the reasons for ED overcrowding. Lee et al. [8] says that the closure of the central hospitals in New York state led to ED overcrowding in other hospitals because ED visits from closed hospitals were redistributed to nearby hospitals. Erik et al. [9] state that to make progress in the health outcomes SDOH factors should be addressed by including them in ED visit dataset. So, a technique will be needed by the hospitals to identify the reasons for ED over-utilization.

In the second problem, we analyze the clinical and social characteristics of Coronavirus Disease 2019 (COVID-19) affected patients in order to identify which populations in LA have an increased risk of getting COVID-19, to support real time care decisions. Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is the virus that caused COVID-19 in the US. The first

COVID case in the US was reported on January 21, 2020 [10]. According to Centers for Disease Control and Prevention (CDC), as of July 8, 2020, 2,982,900 cases are reported across the US with 131,065 deaths [11]. Retrospective studies on COVID-19 cases showed an increased severity of illness in the elderly population of age greater than 60 [12,13]. As per CDC, population in the age category of 85 and more have been hospitalized 13 times more when compared to 18-29 years age group [14]. Additionally, research studies indicate that underlying medical conditions have a greater risk for severe illness with COVID-19 disease [15-18]. As per CDC, few of the underlying medical conditions that are responsible for severe illness are Cancer, Chronic Kidney Disease, Heart Conditions, Obesity, Smoking, Immunocompromised state, Chronic Obstructive Pulmonary Disease [14].

The SDOH factors such as crowded housing, race, poverty level, ethnicity, physical environment (e.g., homelessness, smoke exposure) can have considerable effect on the COVID-19 health outcomes (test result('Positive/Negative'), Living _Status ('Alive/Deceased') [19]. The effect of COVID-19 is high when the social inequalities are associated with underlying medical conditions [20]. LA experienced increase in the Covid cases between March 31-April 7 due to Mardi Grass celebrations occurred on February 25th in New Orleans city. According to CDC, as of July 8, 2020, LA state reported 68,263 cases with a total of 3,319 deaths [11]. Among 64 parishes in LA, Jefferson has 15% of total cases in LA, Orleans reported 12%, and East Baton Rouge has 9.1% [21]. These parishes have highest number of cases when compared to other parishes in LA [21]. Hence it is necessary to identify the characteristics of population in LA who are at higher risk of acquiring the disease.

According to the World Health Organization (WHO) [22], "the SDOH are the conditions in which people are born, grow, live, work and age." Padarthy et al. [23] analyzed other studies and drew conclusions such as 60% of the persons' health status is determined by behavior and social factors such as socioeconomic status (SES), employment, food insecurity, education, community cohesion and more, 10% by clinical care, and 30% by their genetics. These counts prove the importance of SDOH factors in the health perspective. Table 1.1 gives a brief overview of the various categories and subcategories on SDOH factors. According to U.S News & World Report, 2019, out of 50 states in the US, rank (1 is First; 50 is Last) of Louisiana (LA) sate is reported as 50 by considering health care (16%), education (15.8%), economy (13.8%), infrastructure (12.9%), opportunity (12.7%), fiscal stability (10.5%), crime & collections (9.9%), natural environment (8.4%) categories out of 50 states in the US [24]. In this scenario, we can see that the SDOH factors like health care, education, economy, crime & collections, natural environment together sum up to 63.9% in ranking of LA. By this we can see that LA is not in a good condition in most of the aspects of SDOH categories. Moreover, Brunt [25] explains that Community Health Records (CHR) should contain data about the communities that include social, physical, and lifestyle determinants of health. By including such data provides an opportunity to improve community health.

This research study will utilize patient data of OLOL RMC, a medical center in Louisiana and affiliated to FMOLHS. Patient data will be integrated with population, Social Vulnerability Index (SVI), SDOH-related factors, and community information, discussed in Section 3.1.3.

Table 1.1. Social Determinants of Health; Source: Healthy People 2020, ODPHP [26]

CATEGORY	SUBCATEGORIES
Economic Stability	Employment Food Insecurity Housing Instability Poverty
Education	Early Childhood Education and Development Enrollment in Higher Education High School Graduation Language and Literacy
Social and Community Context	Civic Participation Discrimination Incarceration Social Cohesion
Health Care	Access to Health Care Access to Primary Care Health Literacy
Neighborhood and Built Environment	Access to Foods that Support Healthy Eating Patterns Crime and Violence Environmental Conditions Quality of Housing

1.2. Problem Statement

ED overcrowding is one of the major issues in the health care industry in the US [27]. People across the US are using ED services for non-emergency medical needs which leads to ED overutilization, huge demand for various medical services, and financial loss [28]. Claire et al. [29], says that chronic conditions are the widespread cause for ED overcrowding. According to AHRQ, in the 2011 annual year, 3.3 million adults were readmitted to the hospital in 30 days in the US and costed to $41.3 billion for the hospitals [30]. Moreover, according to the representatives of OLOL RMC ED overutilization is also a major issue in EBRP and EDs of FMOLHS faces many problems with ED patients visiting for non-emergency medical needs. Also, closure of hospitals is one of the reasons for ED overcrowding [8]. For example, in consultation with representatives of OLOL RMC the closure of medical centers namely Earl K. Long Hospital/ED, Baton Rouge General Mid-City ED, and Champion Medical Center between 2013 and 2017 changed the EBRP health care landscape leaving the northern part of the city without close emergency health care services or after-hour urgent care clinics such as UC. This led to the overutilization of other health care providers including FMOLHS ED. Baton Rouge General's Mid City ED experienced large financial losses as most of the population visited the ED more due to the closure of Earl K. Long Medical Center, which is one of the major reasons for its closure [31]. To overcome the gap in emergency health care services in the parish and enable closer health care facilities to the residents, OLOL North Baton Rouge ED (OLOL NBR ED) was inaugurated on November 15, 2017. Also, according to Louisiana Department of Health (LDH), COVID-19 patients death rate is high in the

population having underlying conditions such as Hypertension, Diabetes, Cardiac Disease, Chronic Kidney Disease, Obesity, and Congestive Heart Failure (CHF) [32]. As per the CDC, the non-Hispanic black persons, Hispanics and Latinos, and American Indians/Alaska Natives, show higher rates of hospitalization or death (approximately 5 times, 4 times, 5 times more respectively) due to COVID-19 when compared to non-Hispanic white [33]. According to National Center for Health Statistics (NCHS), Table 1.2 shows the comparison of COVID-19 deaths in LA to total deaths in US the by Age and Sex [34].

Table 1.2. Comparison of COVID Deaths by Age and Sex [as of July 8, 2020]

Age and Sex Characteristics	Male (%)		Female (%)	
	US Total:60374	LA Total:1388	US Total:52241	LA Total:1218
Under-1 to 14	0.03	0	0.02	0
15-24	0.80	0	0.10	0
25-44	3.18	3.10	1.58	3.28
45-64	21.09	22.19	11.91	16.42
65 and Over	74.90	74.71	86.40	80.30

In Table 1.2, we can see that approximately 2.33 % of total US COVID deaths are reported in LA [34]. Moreover, the population with the age category of 65 and above contribute around 77.11% of total COVID deaths in LA [34]. The male gender death rate is 6.54 % higher than female death rate in LA [34]. Hence, it is essential to understand and study the clinical characteristics and SDOH factors that are causing an increase in the risk of the LA population to acquire COVID-19 disease with the help of COVID patient data available in FMOLHS. The other research focus is to predict ED high utilizers in advance whenever there is a hospital closure/opening, new services initiation with the help of EHR maintained by all the FMOLHS affiliated hospitals and clinics including OLOL RMC in order to prevent ED overcrowding and financial loss by providing valuable care, primary care, services to the people in a geographical area that has frequent use of emergency and medical services.

Several research studies have focused on predicting ED visits as an ordinal variable (High, Medium, Low) considering multiple ED locations with the help of machine learning models such as multivariate logistic regression, decision trees (DT), and boosted decision trees such as Adaptive Boosting (AdaBoost) [35,36]. Table 1.3 provides details such as task, multiple ED locations, existing facilities, inputs, methods used and output of some recent studies that developed predictive models in healthcare to prevent the ED utilization problem.

Models in Table 1.3 take data from past years available in EHRs and predict ED utilization for the coming years. Moreover, ED prediction outcome of the models finds strong correlation with all predictor variables except for distance present in EHR data for frequent ED use [35,36]. Pereira et al. [36] says that approaching the objective as a regression problem is equally meaningful because the model developed could predict the ED utilization rate as a real value instead of an ordinal variable (High/Low/Medium).

Additionally, the EHRs structured data can contain selected lifestyle and social domains such as race, ethnicity, preferred language, alcohol drinking, and smoking status to predict the ED utilizations rates [37]. Also, unstructured data (e.g., free-text clinical notes) contain information on selected environmental and social domains such as housing issue, social connection/isolation,

and income/financial resource strain [37]. This research study strengthens to use machine learning techniques and routinely available EHR data to develop the predictive model. Also, focuses on including significant SDOH factors and other factors such as fire rates, locations of food stores (grocery stores and convenience stores), health care facilities (ED, clinics, primary care, and UC services), and pharmacies into the EHRs' for identifying the percentage of utilization of services provided in medical care.

Table 1.3. Summary of Predictive Models for ED Prediction

Study Characteristics	ED Utilization Likelihood in Indiana [35]	ED Utilization Likelihood in California [36]	ED Utilization Rate [38]
Consider Multiple ED?	Yes	Yes	No
Existing Facilities only?	Yes	Yes	Yes
Inputs Used	Age, Sex, Visits, Chief Complaints, Distance between hospital and patient's home	Frequencies (Total ED visits, Total Hospital admissions) per patient, Gender, Age, Race, Distance between hospital and patient's home, Medical (sev0, sev1 and comorbidity features)	Gender, Chronic conditions, Patient hospital-visit count, total cost, ED visit count, cost & count
Machine Learning Methods Used	Multivariate Logistic Regression	Logistic Regression (LR), DT, AdaBoost	XGBoost, Recurrent Neural Network (RNN)
Model Output	Frequent vs low ED Users Classification	Future ED Visits for each patient (High/Medium/Low)	ED Visit count as a real value
Limitations	1. SDOH factors were not included 2.Didn't include patients who did not use ED 3.EHR data from other sites cannot use the developed model	1. SDOH factors were not included 2.EHR data from other sites cannot use the developed model	

1.3. Research Objectives

The specific goals for this research work include:

- Develop and validate a machine learning / neural network model for predictive classification of ED high utilizers as (Yes/No) to multiple ED facilities available in LA and identify correlations between the SDOH factors and the other independent variables influencing the utilization of FMOLHS ED.

- Perform a retrospective study and spatial analysis to identify which patient demographics, clinical characteristics and SDOH factors are responsible for increasing the risk of general population acquiring or dying due to COVID-19.
- Develop and document procedures by which FMOLHS and LSU HSC may utilize predictive modeling in future studies.

Chapter 2. Literature Review

2.1. ED Utilization Predictive Models

Machine learning models are used by researchers to predict the medical services utilized by the patients in advance to reduce the major issues such as ED overutilization and financial loss [39]. Various machine learning techniques could be used to predict the ED high utilizers and prevent the closure of the hospitals due to ED overcrowding [40]. Predictive modeling allows the hospitals for proper allocation of resources and could prevent the over usage of resources [36].

Vest et al. [41], performed a case study on ED revisit prediction for a 30-day time period using two-class boosted decision tree algorithms. The outcome was a revisit to ED within 2, 7, 14 or 30 days. The data sample includes data from an urban safety-net hospital across a state in the US. In this study ED revisits are predicted by including 5 different classes of information i.e., 5 different models including 1) SDOH measures only; 2) Current EHR information only; 3) current and historical EHR information; 4) Health Information Exchange (HIE) information; 5) all available information. Table 2.2 gives the details about each class. Moreover, Comorbidity Index Score is used to categorize comorbidities of patients based on International Classification of Diseases (ICD) codes. For each model, five-fold cross validation is used by dividing the data into training sample (80%) and test sample (20%). Area Under Curve (AUC) score, Accuracy, Precision, Recall and F1 score are used to evaluate the models by the classes of information. Table 2.1 gives details about evaluation metrics. True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN).

Table 2.1. Evaluation Metrics for classification tasks [42]

Metric	Explanation	Range
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	[0-1], 1- is perfect classifier
Recall or Sensitivity	$\frac{TP}{TP + FN}$	[0-1], 1- No missing TP
Precision or Positive Predictive Value (PPV)	$\frac{TP}{TP + FP}$	[0-1], 1-No FP
Specificity	$\frac{TN}{TN + FP}$	[0-1], 1- No FN
F1 Score	$2 * \frac{(Precision * Recall)}{Precision + Recall}$	[0-1], 1- is perfect classifier
AUC	Area under curve of sensitivity vs specificity	[0-1], 1- best performance

In Table 2.2, the model using all classes of information was the best model for predicting ED revisits because of highest AUC score around 74 % and F1 score around 49%. But the model with only SDOH information has least AUC score around 61% and F1 score around 7% when compared to other models in this case study. The ED revisits are more among racial and ethnic minorities.

Also, the ED was revisited mostly for non-emergency reasons, high chronic diseases, low SES and higher crime rates. The models shown in Table 2.2 used limited number of potential measures (e.g. primary care, clinical experience, health literacy are not used) within each information class. Also, the results of the model cannot be generalized to non-urban population.

Table 2.2. Summary of Different Models Information Used by Vest et al. [41]

Information Category	Models			
	EHR (Current)	**EHR (Historic)**	**HIE**	**SDOH (Census Tract)**
Demographics	Age, Gender, Ethnicity	None		Median Household Income Below 185% poverty level Total unemployment
Built Environment	None	None		Property crime rate Parcels within ¼ mile of an active park
Timing	AM/PM and weekday/weekend	None		
Condition History	None	Asthma CHF Diabetes Depression Substance abuse history		None
Utilization Services	Non-emergency Injury Alcohol Substance Abuse	Comorbidity index score(mean) 0-no comorbidities higher score- increase use of resources or result to mortality		Asthma prevalence CHF Diabetes prevalence Depression prevalence
		Prior 30-day outpatient visits, hospitalizations same institution(mean)	Prior 30-day outpatient visits, hospitalizations any institution(mean) No of unique prescription in the past 30 days	
Social Circumstances	None			Racial discrimination Births where mother has less than 12 years of education
Behaviors	None			Tobacco prevalence Substance abuse prevalence

Jianmin et al. [35] performed a case study by developing a multivariate logistic regression model for predicting frequent use of ED care for the years 2009 and 2010 from the available electronic registration data of 96 institutions in the state of Indiana, US, from 2008. The model was developed by including age, sex, visits in 2008, chief complaints with 11 different categories such as Respiratory, Gastrointestinal, Neurologic, Skin, Undifferentiated Infection (UDI), Lymphatic, Influenza-like illness (ILI), Dental, Pain, Musculoskeletal, Alcohol, and zip code centroid straight-line distances between patient's location and hospital facility. The outcome was a binary variable (frequent vs low use classification of ED users). In this case study 9 multivariate logistic regression models were developed with cut off for frequent ED users starting from 8 to 16 ED visits. The AUC score, Positive Predictive Value (PPV), sensitivity, and specificity are used in evaluating the developed multivariate logistic regression models. In this study, 2.8 million patients in the overall 3-year study period generated 7.4 million ED visits. The model with >=8 visits has AUC score around 84%, F1 score around 36% and the model with >= 16 visits has AUC score around 92%, F1 score around 35%. The AUC score increased with the increase in the threshold of ED visits i.e. from 8,9,10 and so on to 16 ED. The age between 25 and 44 years, female gender, pain, musculoskeletal, respiratory, and gastrointestinal chief complaints contributed more to the ED visits. Moreover, alcohol accounted for 0.5% of total ED visits. The model developed does not include population who did not use ED, patients' SES data, and EHR data from other sites cannot use the model. Hence, the model is overrated.

Pereira et al. [36] also performed a case study for the prediction of future frequent ED users as a classification task with ordinal values such as high, medium, and low in California state. The research was done with decision trees (DT), boosted decision trees (AdaBoost) and logistic regression (LR) supervised machine learning models by using discharge data records from California-licensed hospitals for the years 2009 and 2010 and evaluated the predictive accuracy for the years 2010-2011, 2011-2012, and 2012-2013. The training and testing on these models is performed such that it predicts whether the patient is a low-frequency user($\leq$ 1 ED visit), medium-frequency (2-4 ED visits), & a high-frequency ($\geq$ 5 ED Visits) by considering different groups of features such as the number of hospital admissions, demographic features such as age, sex, race, the distance between patient and hospital location such as dist1$\leq$5 , 5<dist2$\leq$20, dist3>20(measured in miles), medical comorbidities(presence of two chronic diseases) of patients and severity such as SEV0: the percentage of ED visits with no presence of (Complication/Comorbidity)CC or MCC (Major Complication/Comorbidity); SEV 1: the percentage of ED visits with presence of CC; SEV 2: the percentage of ED visits with presence of MCC; are included. ICD9 diagnoses codes in the EHR records are mapped to the comorbidity features in Table 2.3. The analysis of the features indicates that including distance is only moderately helpful in predicting ED visits. Moreover, including the past EHR data, demographic information, severity, and comorbidity features is very beneficial. As per the evaluation metrics (AUC, precision, sensitivity), the results of three models (LR, DT, and AdaBoost) were consistent across the three years for three classes (low, medium, and high). But the LR model showed higher AUC scores for all three classes with very minimal difference when compared to other models. Moreover, according to this case study predicting moderate ED users is difficult when compared to low and high ED users. Also, in this research paper, it was stated that approaching the objective as a regression problem is equally meaningful.

Table 2.3. Comorbidity Features [36]

Comorbidities	Name	ICD9 codes [43]
AIDS	HIV and AIDS	V08 and 042
ALCOHOL	Alcohol Abuse	305.0
ANEMDEF	Deficiency anemia	281.0 -281.9
ARTH	Rheumatoid arthritis	714
BLDLOSS	Blood loss anemia	280.0
CHF	Congestive heart failure	428.0 – 428.9
CHRNLUNG	Chronic pulmonary disease	416
COAG	Coagulation deficiency	286.7
DEPRESS	Depression	311
DM	Diabetes w/o chronic complications	249.0, 250.0
DMCX	Diabetes w/ chronic complications	249.1-249.9, 250.1-259.9
DRUG	Drug abuse	305.00-305.93
HTN	Hypertension, uncomplicated	401.0 – 401.9 401 -405
HTNCX	Hypertension, complicated	401.0 – 401.9 401-405
Hypothy	Hypothyroidism	243, 244
LIVER	Liver disease	571
LYMPH	Lymphoma	202.0-202.9
LYTES	Fluid and electrolyte disorders	276
METS	Metastatic Cancer	197.0
NEURO	Other Neurological	IV80.09
OBESE	Obesity	278
PARA	Paralysis	342.0-342.9, 344.0-344.9
PERIVASC	Peripheral vascular disorder	443
PSYCH	Psychoses	290-299
PULMCIRC	Pulmonary circulation disorder	415-417
RENFAIL	Renal Failure	584.5, 586
TUMOR	Solid tumor without metastasis	200
ULCER	Chronic peptic ulcer disease	IV12.71, 533.0-533.9
UNCLASS	don’t correspond to any comorbidity	NA
VALVE	Valvular disease	746.83
WGHTLOSS	Weight loss	783.2

Additionally, there are many papers [44,45] on predicting ED visits, UC, ED use among children with epilepsy, etc. and so on. All of these are using supervised machine learning with classification. The evaluation metrics are AUC, PPV, and Sensitivity. On further research on

applying machine learning models as a regression problem will enable us to predict the ED utilization rate as a continuous value instead of classification labels.

In a research paper, Qiao et al. [38] found that non-linear models capture better performance than linear models in prediction of ED visits based on EHR data. Non-linear models such as XGBoost, Recurrent Neural network (RNN) have a complex model structure and strong fitting ability. So, these models are used to solve the problem of making better predictions when complex correlations are involved [38]. In this paper, both linear and non-linear models are used to predict two tasks. Task 1 is predicting ER Visit count (a real value) as regression and task 2 is predicting ED Visit (Yes or No) as classification. The features included are gender, age, chronic conditions, patient-hospital visit count and total cost, ER visit count, cost & count. The feature dataset contains 6 thousand patients, 0.1 million visits and hundreds of unique diagnose codes from a single hospital. The validation set, training set, and testing set are divided in the proportion of 1:8:1. Finally, the evaluation of the models is done through R-squared value (It is a statistical measure to identify how close the data are to the fitted regression line) as part of regression evaluation metrics. Also, AUC (measure of quality of classification models) value as part of classification evaluation metrics. So, for task 1 RNN model gave good results with R-Squared value around 34% whereas for task 2 XGBoost model gave good results with AUC value around 70%.

2.2. Healthcare Prediction Models

There are research papers on using prediction models in the healthcare such as predicting the patient length of stay (LOS) is determined using Artificial Neural Networks (ANN) by Muhammet et al. [46]. Here, patient LOS is predicted using input factors such as patient age, sex, mode of arrival (walk-in patients (foot, special car, stretcher), ambulance vehicle (car, helicopter, etc..), ambulance patients), treatment unit (5 different areas based on acuity level), medical tests and inspection in an emergency department. The dataset is taken from a single hospital emergency department in the eastern part of Turkey. A total data of 1500 ED patients who were treated in the department in October and November 2010 are considered. The correlation between the variables is conducted to identify strong, weak, and medium correlations with a test probability value(P-value). Different learning rates, momentums are used with various learning algorithms such as Quasi-Newton, Online-Back Propagation and Levenberg-Marquardt [46]. The best fit model is chosen with the lowest absolute error. By predicting accurate patient LOS enables the ED management to have correct resource allocations and appropriate utilization of the resources. But this paper shows an R-squared value of 63% for ANN model with Levenberg-Marquardt training algorithm. This is not an ideal prediction as it is less than 80%. So, other techniques such as logistic regression, support vector machine, and multiple linear regression are suggested [46].

Choi et al. [39], performed a case study on early detection of heart failure using data from a health system. The RNN, logistic regression, support vector machine, k-nearest neighbor machine learning models are used to predict the heart failure based on the input variables. The input data set has demographics, tobacco and alcohol consumption, clinical and laboratory values, ICD-9 codes, and prescription information. Finally, the RNN model has showed good AUC score around 88% for 18-month observation window when compared to other models in this case study.

Benjamin et al. [47], says that six machine learning models are used to predict systolic blood pressure, body mass index, waist circumference, and telomere length. They are two Ordinary Least

Squares (OLS) regression models (minimal, theory-based). The other four models are linear regression, penalized regressions, random forests, neural networks. The data is from Health and Retirement Study (HRS). Also, the R-squared value was considered to estimate the best fit model. In this paper, neural networks are the best fit model with R-squared value between 40-60% whereas for other models it is less than 30%.

As per the literature review of various researchers, the major findings are approaching the research objective as a regression task. Also, using linear models such as regression analysis and nonlinear machine learning models such as neural networks for complex relationships. However, as per the literature review the SDOH factors were not included and the developed models were not used in different locations. Hence, in this research study, the population and social determinant of health (SDOH) factors influencing utilization will be included in the model development. Therefore, it strengthens to use machine learning models such as neural networks, multivariate regression. So, prediction of ED high utilizers, risk of readmission within 30 days for CHF, and correlated factors in using the medical services will be done using machine learning techniques and prevent the ED overcrowding by providing valuable care to the patients in order to reduce frequent visits to ED.

2.3. SDOH Factors Affecting ED Utilization

Andrieni [48] says that "Social determinants of health affect every aspect of healthcare. We're learning that we have to address not only clinical factors that affect the patient's health but also the nonclinical ones if we are going to successfully treat patients." Johns Hopkins University (JHU) researchers identified lack of insurance, lack of primary care physician, psychiatric illness, substance abuse, cognitive/physical impairment and difficulty understanding discharge instructions as risk factors associated with high ED use among three hospitals in the US [49].

Various studies have mentioned that population who are about to become homeless or already homeless utilize ED more when compared to the general population [50,51]. For example, Hwang et al. [52] found that the homeless population visits ED 8 times more than the non-homeless counterparts. Doran et al. [53] mentioned that the public uses ED, not for health crisis alone but also during their life crises. Also, the author conducted a study on a cross-sectional data sample of 625 ED patients, from a single hospital center. Table 2.4 compares the study results with the general population in the US.

Table 2.4. Comparing ED Patients versus General Population

SDOH Factors	% of people visiting ED Total:625 ED Patients [20]	General Population US Pop.: 325.7 million
Food Insecurity	40%	11.1% [54]
Homeless	34%	0.17% [55]
Didn't Meet Essential Expenses	42%	18% [56]
Money Concerns	28%	7% [56]

Moreover, Heath et al [57], mentioned that experiencing a social barrier led to more healthcare utilization i.e. patients with food insecurity visited ED more than two times in 12 months when compared to non-food insecurity people. Also, patients who reported public safety issues are admitted to ED 3.2 times more than the people without public safety issues [57]. Finally, the outcome was that the high rates of social needs are among people, including homelessness, food insecurity, and inability to afford medications and basic needs [53]. According to Axelson et al. [58], food insecurity, low literacy, economic insecurity, housing and homelessness and neighborhood & access to safety are the main SDOH factors that are responsible for frequent ED utilization.

Surveys play the main role in identifying the SDOH factors that are responsible for frequent ED visits [59]. The outcome measures of the survey are age, gender, ethnicity, employment, housing status, insurance, access to food and the choice between buying food and medicine. The conclusion was most of the ED patients have food insecurity and hunger when compared to other factors of SDOH. Griffey et al. [60] analyzed that health literacy (capacity to obtain, process, and understand basic health information and services to make correct health decisions) plays the main role in ED visits and concluded that the increasing order of ED visits is inadequate health literacy> marginal health literacy> adequate health literacy. Runyan, C.N, [61], informed that the National Academy of Medicine recommended 12 SDOH domains into the electronic health records as part of primary care. They are ethnicity/race, tobacco use, alcohol use, residential address, educational attainment, financial resources, stress, depression, physical activity, social isolation, intimate partner violence, and neighborhood median household income.

Boston Medical Center [62] developed a social needs screener as part of EHR which captures SDOH data to have an improvement in the outcome of patients. Also, Fenton [63] stated that 85% of physicians believe patients social needs are as important as their medical conditions. As per the special report by Sukel [48] a study was performed by Kash and Colleagues on SDOH. In the report, the predictive model techniques were used to determine which SDOH factors are most helpful to predict avoidable hospital readmission by using EHRs' data and data from the area deprivation index (ADI)—census data that looks at common socioeconomic factors. Therefore, SDOH factors are influencing an individual decision to visit the ED for non-urgent conditions [28].

2.4. Retrospective Studies of COVID-19

Price-Haywood, E.G., et al [64] conducted a retrospective study on the COVID-19 positive tested patient data of Ochsner Health in LA between March 1 and April,11, 2020. This study included demographic characteristics such as age, sex, race, ethnic group, and insurance plan); chronic conditions associated to ICD-10 codes (E66, Z72.0, J45, J44, E10, I10, I50, I25, N18, Z94, K70); other clinical characteristics such as Body Mass Index (BMI, calculated by dividing weight by height); outpatient medications such as immunosuppressants, glucocorticoids, chemotherapy, and immune modulators; diagnosis codes related to primary care, Urgent care and ED during COVID-19 testing; diagnosis codes linked to inpatient encounters; Zip codes to identify low-income areas. In this case study, 3481 COVID-19 positive patients are included in which 60% are female, 70.4% are black non-Hispanic, 29.6% are white non-Hispanic.

A multivariable analysis was performed, and the results showed that black ethnicity Covid patients, increasing age, high Charlson Comorbidity Index (higher score implies less healthy),

public insurance (Medicaid or Medicare), residence in low-income area, obesity factors are highly correlated with increase of hospital admissions [64]. Though the black ethnicity Covid patients showed a higher percentage of COVID deaths, the blacks ethnicity comprises only 31% of Ochsner Health's patient population. Price-Haywood, E.G., et al [64] mentioned that black ethnicity Covid patients was not associated with higher in-hospital mortality than white ethnicity Covid patients in their health care system.

Kim, S.J.; et al [65] has done spatial analysis of COVID-19 deaths for identifying spatial clusters of social vulnerability and health risk factors in Chicago. In this study, the author performed spatial autocorrelation for percentage of African American COVID-19 death rate. The Chicago state has 268 deaths and in those 62.8% are African Americans. In this case study, it was mentioned that African American communities are affected by multiple chronic diseases before the COVID-19 pandemic and due to which they have high mortality rate in Chicago. The spatial analysis showed that increase in social vulnerability is highly associated with health inequity.

A research study [66] mentioned the importance of including SDOH data with the COVID-19 data. The socioeconomic position (SEP) is largely related with the COVID disease and mortality. SEP plays a vital role in the COVID-19 outbreak either directly or indirectly through occupation, living conditions, health related behaviors, presence of comorbidities and immune functioning. For example, a person who has the occupation that requires to interact with many people such as retail staff, cleaners, healthcare workers, living conditions such as low-income, crowded housing and, low education will be associated with many risk factors that might increase the risk of acquiring COVID-19. Though we have studies on the importance of SEP, the influence of these factors on COVID-19 transmission from person to person, severity, and outcomes is not known. Hence, we are integrating FMOLHS COVID-19 patient data with the SDOH data to identify the factors that are influencing the spread of the COVID-19 disease.

Chapter 3. Methodology

A retrospective study is performed by taking data from various platforms and medical records of FMOLHS to gather the datasets of patients in LA. The collected patient data from EHR of FMOLHS is integrated with population and SDOH-related factors available through public sources such as census data discussed in Section 3.1.3. The ED high utilizers classification predictive model is developed with the help of machine learning techniques. Also, the COVID-19 tested patients clinical characteristics and social factors that are responsible for the increase in the spread of the disease are analyzed spatially using ArcGIS. Further statistical analysis was performed to identify correlations between the various factors including SDOH and the other independent variables influencing the utilization of FMOLHS ED and responsible for the FMOLHS COVID-19 deaths. The detailed description of data collection, preparation, analysis and methodology are explained in this chapter.

3.1. Data Collection and Preparation

A common EHR is maintained by all the hospitals including OLOL RMC affiliated to FMOLHS for storing patient's data and sensitive information. The EHRs of the hospital contains all the EMR of the patients and Protected Health Information (PHI) in a password protected excel sheet linked to Study Identification Number (SIN) such as Address, Date of Birth (DOB), Full Name, and Medical Record Number (MRN). An electronic search of the EHR data is performed using Epic and Cerner software's in FMOLHS to get the needful patient data. Using the electronic search capability, all the ED visit information is obtained according to the filter applied. This information contains patient demographic details such as name, medical record number, date of birth, age, sex, ethnicity, payer type, and address. Also, includes clinical information such as date of ED visit, ICD9 and ICD10 codes, and disposition of ED visit (admit or no admit). Additionally, an electronic search is performed to get information related to patients diagnosed with COVID-19. This dataset contains all the demographic and clinical information mentioned above along with type of visit (Inpatient or outpatient). Women and minorities are included in the datasets. Adult patient and child patient data is included in the ED visit information but the patients under the age of 18 are excluded from COVID-19 dataset.

In this research study data is collected from various platforms and primary portion of the data is collected from EHR of FMOLHS, census data and community information available through US Census Bureau, American Community Survey (ACS), publicly available community information like shape files, and health related geographic data. This research is approved by Franciscan Missionaries of our Lady University (FMOLU) with IRB number 10102 approved on 06/28/2018, LSUHSC with IRB number 2018-151 approved on 10/04/2018 and, LSU with Institutional Review Board (IRB) number 4299 approved on 10/24/2019. The IRB approval letters From OLOL RMC, FMOLU, LSUHSC, LSU are shown in Appendix A, Appendix B, Appendix C, and Appendix D respectively. The waiver letter from Health Insurance Portability and Accountability Act of 1996 (HIPAA) Authorization from LSUHSC is shown in Appendix E.

3.1.1. ED Patient Data from EHR of FMOLHS

In this research study patient data related to ED visits affiliated with FMOLHS were obtained from the EHR for ED visits between January 2015 and January 2020. This ED data contains the

information of patients visited ED at least once between January 2015 and January 2020. Table 3.1 shows the information contained in a single row of ED EHR dataset.

MRN is a unique identification number. This number is specific to each organization such as FMOLHS has its own format for the number. It is used to store and get details of patient medical records. Also, the purpose of ICD is to identify health related conditions with a common language at any health care organization.

Table 3.1. Information in Each ED Patient Record

Category	Details
Admission	Date of admission Date of discharge time of visit Health center visited Length of Stay Admit Status
PHI	MRN First name Last Name Date of Birth Address
Diagnosis	ICD 9 and ICD 10 Codes Description Group
Other	Race Gender Financial Class (Medicaid, insurance coverage etc..)

3.1.2. COVID Patient Data from EHR of FMOLHS

Data on patients tested for SARS-CoV-2 between March 13 and June 15, 2020 across the FMOLHS affiliated hospitals and clinics (include LPG, LSU Health and Baton Rouge Clinic) was obtained from the EHR. Table 3.2 shows the information contained in a single row of this ED EHR dataset.

Table 3.2. Information in Each COVID Patient Record

Category	Details
Admission	Hospital Unit Hospital Admission and Discharge Dates Ordering Department, Ordering Date Lab, Collection Date, Admit Type Test (SARS-CoV-2 PCR, SARS-CoV-2 RNA) Result (Positive, Negative)
PHI	MRN, First name Last Name, Date of Birth, Address
Other	Race, Ethnicity, Gender, BMI (Height/Weight), Discharge Disposition, Living (Alive, Deceased)

3.1.3. Census Data and Social Vulnerability Data

Census data, community information, social vulnerability are obtained through various online sources. This information is imported into the GIS as geospatial layers. The brief explanation of community information and context measures that were included and their respective online sources are shown below:

Community information: Community information is publicly available through online sources such as US Census data and the American Community Survey (ACS). Demographic composition, education level, employment sectors, occupations, food, household composition, housing conditions and vehicle access across census tract, census block group and census block level are considered as part of community information. Additionally, information related to census tracts, zip codes, census block groups, and census blocks are publicly available through US Census Topographically Integrated Geographic Encoding and Referencing (TIGER) in the form of shape files. These shape files are easily imported into the GIS system as layers.

Social Vulnerability Data: According to CDC, Social Vulnerability is defined as the degree to which a community exhibits certain social conditions, including high poverty, low percentage of vehicle access, or crowded households, may affect that community's ability to prevent human suffering and financial loss in the event of disaster [67]. The Agency for Toxic Substance and Disease Registry's (ATSDR's) Geospatial Research, Analysis & Services Program (GRASP) created Centers for Disease Control and Prevention Social Vulnerability Index (CDC SVI or SVI). This SVI is useful to what extent and in which social factors the U.S census tracts are vulnerable and would need support before, during, and after a hazardous event or disease outbreak [67]. In this research study, we have included the SVI data for LA downloaded from CDC'S SVI Data as a shape file with the help of ArcGIS. Also, the sources that are used to get each variable and their descriptions are available in the SVI Documentation PDF.

3.1.3.1. SVI Data to FMOLHS Patient data

Table 3.3 shows the CDC'S SVI Data included in the research study. This data is collected at census tract level of LA. For each field calculation and specific table names of data refer to the SVI Documentation PDF.

3.1.3.2. ACS Data to FMOLHS Patient data

Table 3.4 shows the American Community Survey (ACS) included in the research study. This data is collected by 2018 ACS 5-Year Estimates Data Profiles at census tract level of LA. Additionally, the Rural-Urban Commuting Area (RUCA) primary (Prim_RUCA) codes that differentiates each US census tracts were integrated to the FMOLHS patient data. These codes are the measures of population density, urbanization, and daily commuting in the US census tracts [68].

3.1.4. Software Required

In this research study, ArcGIS Desktop (Version 10.6.1) a GIS tool is used to map the patient's data (geocoded addresses), census data, community information, and SDOH factors available from various online public sources such as CDC'S SVI Data and American Community Survey. Also, SPSS software (Version 26) is used for statistical analysis purposes, i.e., to identify the correlations

between the independent variables (inputs provided) and the outcomes. Python 3.7 version is used to develop the prediction model by using Random Forests, Regular Gradient Boost, XGBoost, Neural Networks machine learning techniques. Also, the needful libraries are imported in python

Table 3.3. SVI Data; Source: CDC SVI 2018 Documentation [67]

Variable (Census Tract)	Description (Numerical Values)
E_TOTPOP	Population estimate, 2014- 2018 ACS
EP_POV	Percentage of persons below poverty estimate
EP_UNEMP	Unemployment Rate estimate
EP_PCI	Per capita income estimate, 2014- 2018 ACS
EP_NOHSDP	Percentage of persons with no high school diploma (age 25+) estimate
EP_AGE65	Percentage of persons aged 65 and older estimate, 2014- 2018 ACS
EP_AGE17	Percentage of persons aged 17 and younger estimate, 2014- 2018 ACS
EP_DISABL	Percentage of civilian noninstitutionalized population with a disability estimate, 2014-2018 ACS
EP_SNGPNT	Percentage of single parent households with children under 18 estimate, 2014- 2018 ACS
EP_MINRTY	Percentage minority (all persons except white, non-Hispanic) estimate, 2014-2018 ACS
EP_LIMENG	Percentage of persons (age 5+) who speak English "less than well" estimate, 2014- 2018 ACS
EP_MUNIT	Percentage of housing in structures with 10 or more units estimate, SVI
EP_MOBILE	Percentage of mobile homes estimate, DP04
EP_CROWD	Percentage of occupied housing units with more people than rooms estimate, SVI & DP04
EP_NOVEH	Percentage of households with no vehicle available estimate, DP04
EP_GROUPQ	Percentage of persons in institutionalized group quarters estimate, 2014- 2018 ACS
RPL_THEME1	Percentile ranking for Socioeconomic theme summary. (Sum of percentile rank of EP_POV, EP_UNEMP, EP_PCI, EP_NOHSD)
RPL_THEME2	Percentile ranking for Household Composition theme summary. (Sum of percentile rank of EP_AGE65, EP_AGE17, EP_DISABL, EP_SNGPNT)
RPL_THEME3	Percentile ranking for Minority Status/Language theme. (Sum of percentile rank of EP_MINRTY, EP_LIMENG)
RPL_THEME4	Percentile ranking for Housing Type/ Transportation theme. (Sum of percentile rank of EP_MUNIT, EP_MOBILE, EP_CROWD, EP_NOVEH)
RPL_THEMES	Overall percentile ranking (Sum of all RPL THEMES)
EP_UNINSUR	Adjunct variable - Percentage uninsured in the total civilian noninstitutionalized population estimate, 2014- 2018 ACS

Table 3.4. ACS Data; Source; Table: DP03 (US Census data) [69]

Variable (Census Tract)	Description (Numerical Values)	DP03 Table Field Names
EP_OCC_MBSA	Percentage of persons (age 16+) with Management, business, science, and arts occupations	DP03_0027PE
EP_OCC_SER	Percentage of persons (age 16+) with Service occupations	DP03_0028PE
EP_OCC_SAL_OFF	Percentage of persons (age 16+) with Sales and Office occupations	DP03_0029PE
EP_OCC_NRCM	Percentage of persons (age 16+) with Natural resources, construction, and maintenance occupations	DP03_0030PE
EP_OCC_PTMM	Percentage of persons with Production, transportation, and material moving occupations	DP03_0031PE

3.1.5. Integrating FMOLHS patient data to Census data

Geocoding is the process of converting any physical address into latitude and longitude coordinates. As part of this research study, Texas A&M University Geoservices [70] are used to geocode the FMOLH's patient's address. By using ArcGIS software, the census data such as census tract information, state, parish name, SVI data, ACS data are loaded into the tool as shape or layer filed. Now, the geocoded patient's data is also loaded into the tool with the help of latitude and longitude coordinates. Spatial Join and Attribute Join techniques are used to map the census related information to the FMOLHS patient data.

ED Visits Dataset: The FMOLHS ED Visit data set contains all the ED visits between EHR between January 2015 and January 2020. The patient visits between January 2015 and January 2016 are mapped with 2016 LA SVI data and 2010 LA Primary RUCA codes whereas the patient visits between January 2017 and January 2020 are mapped with 2018 LA SVI data and 2010 LA Primary RUCA codes.

COVID-19 Dataset: The FMOLHS COVID-19 data set contains the COVID-19 disease patients' information between March 13 and May 6, 2020. This data is mapped to 2018 LA SVI data, 2018 ACS 5-Year Estimates DP03 data Profiles and 2010 LA Primary RUCA codes.
The Attribute tables of these layers are exported to csv/excel files. These files can be imported into SPSS software and save as .sav files to perform the statistical analysis.

3.1.6. FMOLHS ED Patient Data

The FMOLHS ED Patients visits data set has more than 839K (839,684) patient visits. As part of this research study, after consulting with FMOLHS staff, we counted the ED visits of a patient (ED_Unnecessary_Visits) with Admit status as ED Discharge, because the visit to ED with Admit status as Admitted is considered to be a valid visit and the patient should get treated in ED. Out of 839K ED Visits between the 2015-2019 years, approximately 545K (545,008) are unique patient visits. These unique patients are filtered based on MRN and year of visit. For example, if the combination of MRN and year exists more than once in a year then duplicate record is removed from the data set and Length of Stay (LOS), ED_Unnecessary_Visits value for each patient in a year is summed in the dataset to maintain a unique record. In this research study, the features that are considered as dependent and independent variables for predicting ED high utilizers are shown in Table 3.5. Additionally, in the Section 3.1.3 the details of census data (SVI and ACS) data sources are explained. Moreover, the specific data field descriptions that is integrated to the ED patient visits data and are shown in Table 3.3 and Table 3.4.

In Table 3.5, the dependent variable ED_Unnecessary_Visits is calculated based on the Admit field. If the Admit field value is ED Discharge then the visit to ED by a patient is counted and considered as an inappropriate visit. Therefore, the ED_Unnecessary_Visits field value is 0 if the Admit field value is Admitted. This implies that the patient visits to ED are appropriate. Moreover, the dependent variable ED_HighUtilizers is derived based on ED_Unnecessary_Visits. If the ED_Unnecessary_Visits field value is 4 or more per year, the patient is treated as ED high utilizer (1) and (0) vice versa. The Independent variables LOS is calculated as the difference in days between hospital discharge and hospital admission dates whereas Age is calculated based on patient's date of birth and hospital admission date. Also, if the MRN and Year of a patient Urgent

Care Visit (UC_Visits) matches with the MRN and year of ED Visit then the MRN is counted and the count per year is assigned to UC_Visits field. The same calculation is done for Primary Care Visits (PC_Visits) per year. In the FMOLHS ED Visits data set we have 391K (391,243) "ED Discharge" records and 153K "Admitted" records. As we are looking at ED_Unnecessary_Visits, we removed the 153K "Admitted" status records from the dataset. Appendix F shows the specific values for the categorical variables included in the feature selection and their frequencies

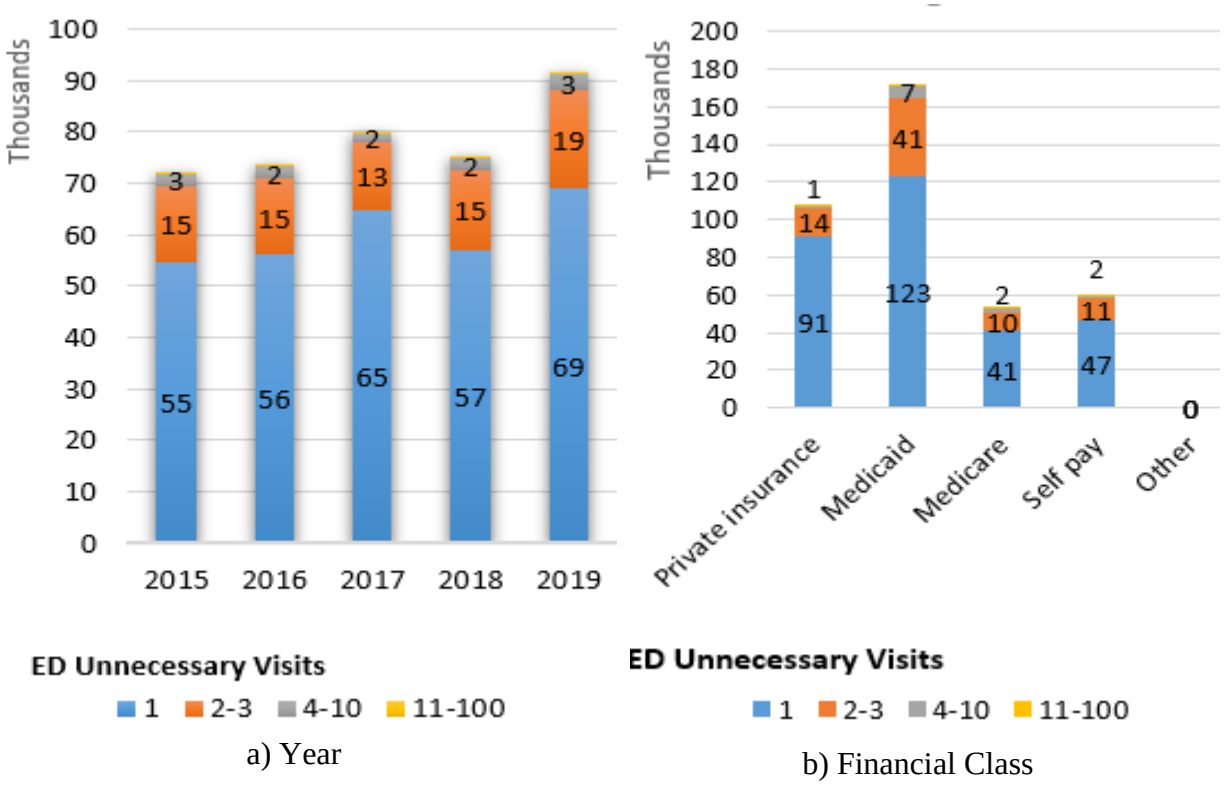
Table 3.5. List of Dependent & Independent Variables for ED Utilization Prediction Model

Factors	Variable Type	Category
ED_Unnecessary_Visits	Numerical	Output/dependent
ED_HighUtilizers	Numerical	
Independent Variables		
Year	Categorical	Patient Data
Length of Stay (LOS)	Numerical	
Admit (ED Discharge, Admitted)	Categorical	
Financial Class (Private insurance, Medicaid, Medicare, Self-pay, Other)	Categorical	
Age	Numerical	
Race (White, Black, Hispanic, Asian, Other or Unknown)	Categorical	
Gender (Male and Female)	Categorical	
PC Visits Per Year (Primary Care)	Numerical	
UC Visits Per Year (Urgent Care)	Numerical	
E_TOTPOP	Numerical	LA_Census Tract_SVI Data
EP_POV	Numerical	
EP_UNEMP	Numerical	
EP_PCI	Numerical	
EP_NOHSDP	Numerical	
EP_AGE65	Numerical	
EP_AGE17	Numerical	
EP_DISABL	Numerical	
EP_SNGPNT	Numerical	
EP_MINRTY	Numerical	
EP_LIMENG	Numerical	
EP_MUNIT	Numerical	
EP_MOBILE	Numerical	
EP_CROWD	Numerical	
EP_NOVEH	Numerical	
EP_GROUPQ	Numerical	
RPL_THEME1	Numerical	
RPL_THEME2	Numerical	
RPL_THEME3	Numerical	
RPL_THEME4	Numerical	
RPL_THEMES	Numerical	
EP_UNINSUR	Numerical	
Prim_RUCA	Categorical	2010 Census Tract Data

Appendix F data shows that Black ethnicity patients visit the ED for unnecessary reasons 1.6% more than Whites. FMOLHS patients between 21-50 age visits ED for unnecessary reasons 38.7% more than patients with age above 60. Also, patients with Medicaid visit ED 16.5 % more than Private Insurance patients. According to year wise percentages we can see that the unnecessary visits to ED have increased from 2015-2017, decreased 1.3% from 2017-2018, and increased 4.2% from 2018-2019. Hence, we integrated the SVI and ACS data to analyze the social factors that are leading to increase in the non- emergency visits to ED and identify the High Utilizers.

3.1.6.1. ED Unnecessary Visits and High Utilizers Data

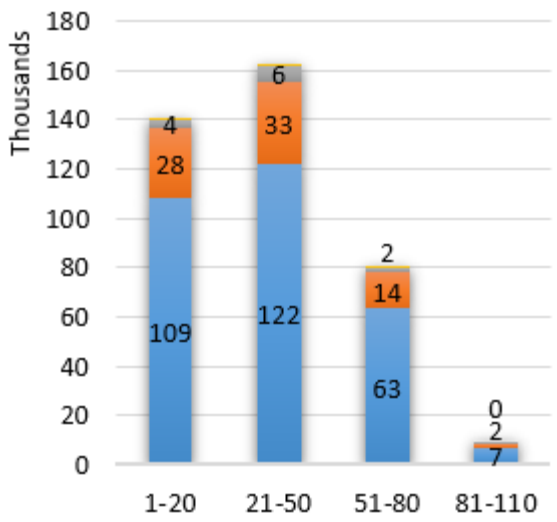
Figure 3.1 shows the distribution of overall ED_Unnecessary_Visits count of FMOLHS ED patients in LA with respect to year, insurance category, age, race, gender, and RUCA codes. From these bar charts we can see that approximately more than 95% of ED unnecessary visits count per year in FMOLHS are between 1 to 3 but more than 70% of these visits are 1. Less than 4 % of the entire ED Unnecessary visits have count between 4 to 10 and very few of the patients (less than 0.5%) have count between 11 to 100.

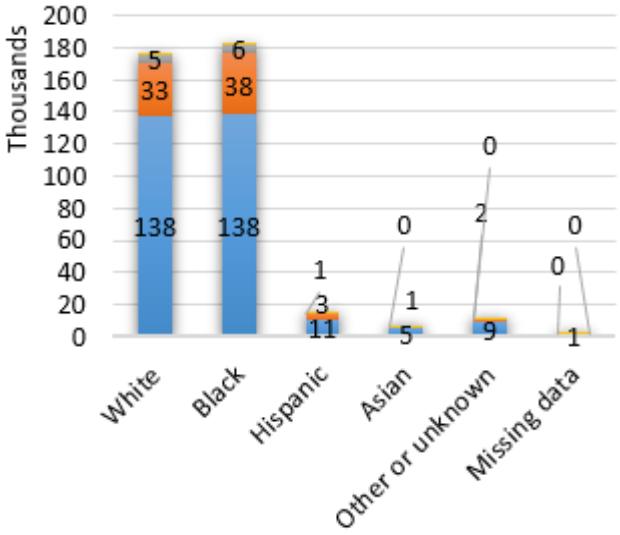


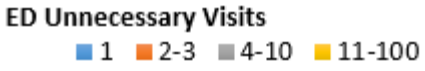
a) Year

b) Financial Class

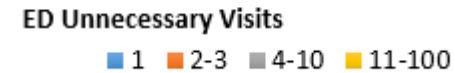
Figure 3.1. LA ED Patient Population Unnecessary Visits Count Distribution With a) Year, b) Financial Class, c) Age, d) Race, e) Gender f) Prim_RUCA



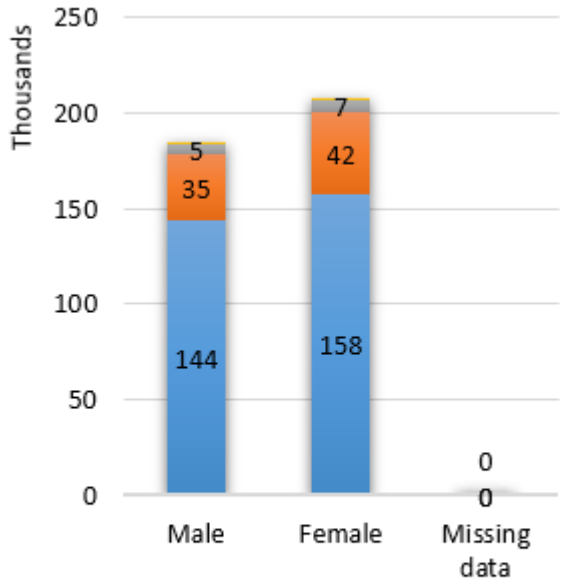


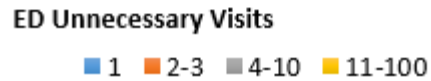


c) Age

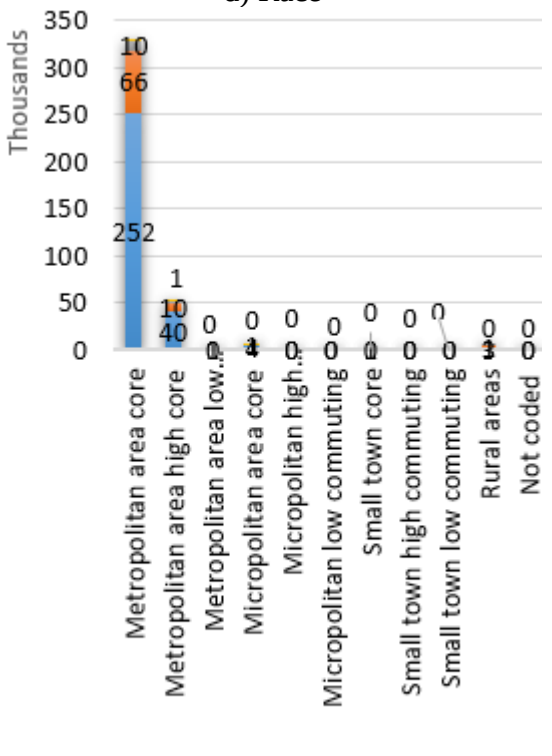


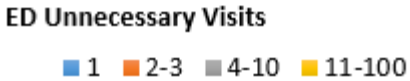
d) Race





e) Gender

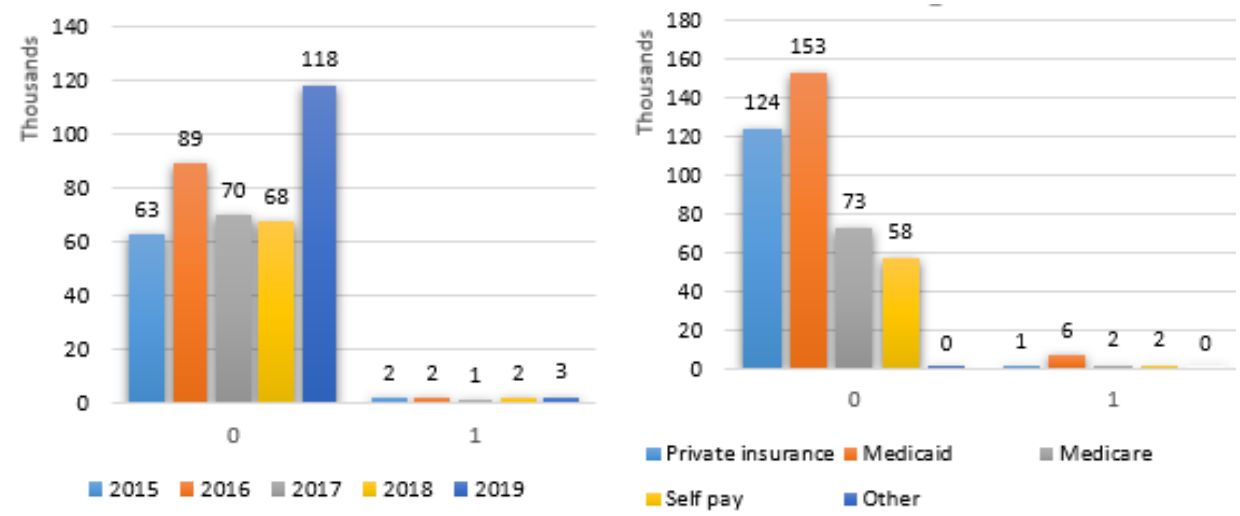




f) Prim_RUCA

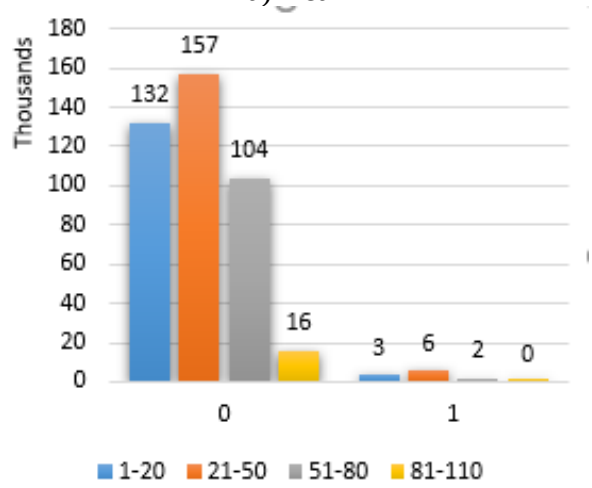
Figure 3.1. (cont'd) LA ED Patient Population Unnecessary Visits Count Distribution With a) Year, b) Financial Class, c) Age, d) Race, e) Gender f) Prim_RUCA

After understanding/studying the year wise FMOLHS ED visits data based on the ED unnecessary visits (calculated per year), we observed that same patient visited the ED across the years 2015-2019 more than once. To study the ED High Utilizers, the data is filtered based on MRN and removed the duplicates across the years. If a patient has visited multiple times across the years, then the latest information of the data is considered for the unique patient record. By removing the duplicates across the years, we now have total of 419,398 unique records across the years. Out of 419,398 records we have 297,154 "ED Discharge records" and 122,244 "Admitted" records. As discussed in earlier sections, we considered a patient as ED High Utilizer based on ED Unnecessary Visits per year. We removed the 122K "Admitted" records. Figure 3.2 shows the distribution of FMOLHS ED High Utilizers count in LA with respect to year, insurance category, age, race, gender and, Primary RUCA. In the FMOLHS ED patient visits dataset, ED unnecessary visits between 1 to 3 are considered as Not-ED High Utilizer (denoted by "0") and 4 -100 ED unnecessary visits are considered as ED High Utilizer (denoted by "1").

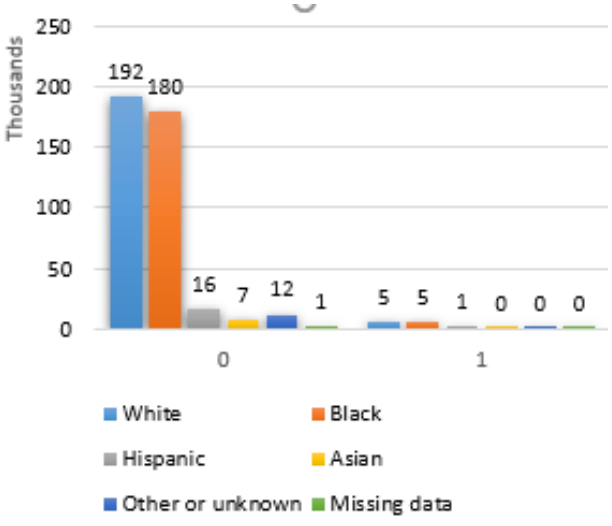


a) Year

b) Financial class

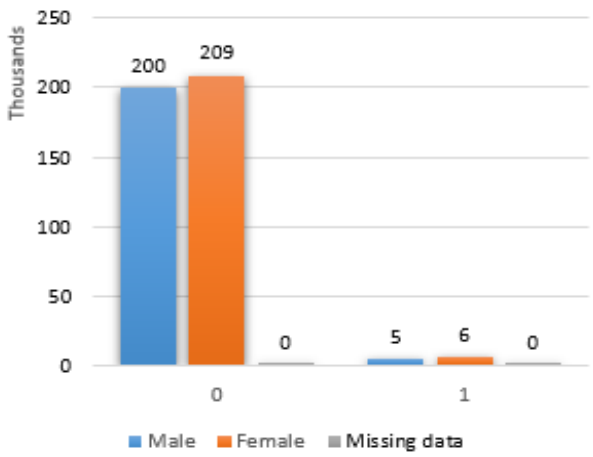


c) Age

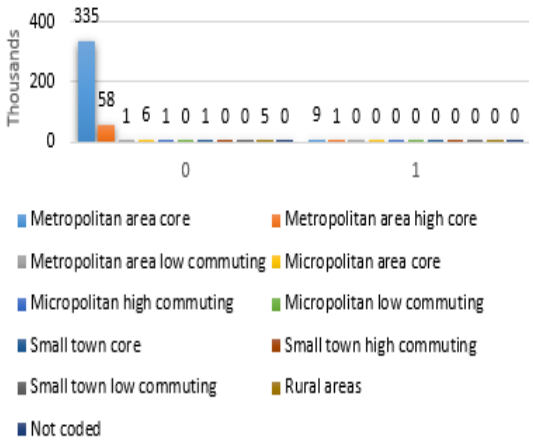


d)Race

Figure 3.2. LA Not-High (0) and High (1) ED Utilizer's Patient Population Visits Count Distribution with a) Year, b) Financial Class, c) Age, d) Race, e) Gender f) Primary RUCA



e) Gender



f) Primary RUCA

Figure 3.2. (cont'd) LA Not-High (0) and High (1) ED Utilizer's Patient Population Visits Count Distribution with a) Year, b) Financial Class, c) Age, d) Race, e) Gender f) Primary RUCA

Out of 297K ED Discharge patients there are 10,996 ED High Utilizers. From the first graph in Figure 3.2 we can observe that among the total ED High utilizers 22.35%, 19.43%, 13.52%, 19.88%, 24.82% are in 2019, 2018, 2017, 2016, and 2015, respectively. In the second graph Medicaid insurance category patients' attributes to 56.68% of all ED high utilizers. Next, the third graph interprets that the patients between the age of 21-50 contributes to 50.41%. Based on fourth graph black ethnicity patients are 49.18% whereas white ethnicity patients are 43.05%. Also, the fifth graph tells that female ED high utilizers are 13.12% more than male patients. The final graph shows that 85.87% of all the ED high utilizers are from metropolitan area core. These statistics indicate that black ethnicity people, female gender, with increasing age, having insurance category as Medicaid and, coming from metropolitan are cores are utilizing the FMOLHS ED more.

3.1.7. COVID Patient Data

The FMOLHS registry of patients tested for SARS-CoV-2 between March 13th, 2020 to June 15th, 2020 SARS-CoV-2 between March 13th, 2020 to June 15th, 2020 contains 18,415 people. Among these tested patients 2,938 people had a positive test result. In this research study, we included the patients with a LA address. In total for LA, 18,083 were tested and 2,902 were positive. After looking at the distribution of the COVID test data we identified that children (anyone under the 18 years of age) are going to present very different from adults. For example, their BMI is much lower, Charlson Predictive Mortality Index (CPMI) indicates the 10-year predictive survival score and is very high when compared to adults. They do not have the comorbidities at the same rate as adults, the percentage of children coming from institutions (nursing homes, rehab, jail, or mental health) is negligible when compared to adults (For example, in FMOLHS COVID dataset we have only one patient record under 18 coming from prison). Hence, for this retrospective study we included a total of 16,472 patients tested for COVID-19, out of which 2,758 were tested positive.

3.1.7.1. Assigning Clinical Data to COVID Patient Data

As discussed in earlier sections that underlying problem history, SDOH factors of an individual might increase the severity and chance of acquiring COVID-19. In this retrospective study, we

have included the 1,214 different problems grouped into 16 categories shown in Table 3.6. Additionally, CPMI is obtained from FMOLHS' Epic software. A score of 0 indicates that the patient survival chance is very less in the next 10 years and the higher score indicates that there is very less of mortality in the next 10 years. In our research study, this score is mapped to the COVID patient's dataset. In addition, the general patient demographics such as age, gender, race, ethnicity, financial class are also available in the dataset.

Table 3.6. Problem List and Associated Category

Problem List	Category
Auto immune diseases	Auto immune
Cerebrovascular Disease	CEVD
Chronic Kidney Disease	CKD
Cirrhosis Liver Disease	CLD
Cardiovascular Disease	CVD
Diabetes	Diabetes
En-Stage Renal Disease	ESRD
Hepatitis B	Hep B
Hepatitis C	Hep C
Huma Immunodeficiency Virus	HIV
Hypertension	HTN
Immunodeficiency diseases	Immunocompromised
BMI, Morbid Obesity related diseases	Obesity
Exceeding BMI range	Overweight
Peripheral vascular disease	PVD
Asthma, Bronchial and, all respiratory diseases	Respiratory

3.1.7.2. Mapping Institutions to COVID Patient Data

Institutionalized patients are the patients who cannot live independently and stay in different institutions such as Nursing Homes, Rehab, Mental Health, Group Home and, Senior Living Homes. Additionally, people who are staying in prison are also considered as institutionalized patients. In this research study, we have people coming from 115 different institutions across LA to FMOLHS affiliated hospital and clinics for the SARS-CoV-2 test. We identified different institutional groups by grouping the people staying in Nursing Homes, Senior Living and, Group Homes as Nursing Home, people from Rehab, Mental Health as Mental Health/Rehab, people from Prison, Jail as Jail/Prison. After grouping, if the address of the patient registered at the time of visit matches with the institution address from the list of 115 institutions then we mapped that particular institution name, institution group (1-Nursing Home, 2-Mental Health/Rehab, 3-Prison\Jail) to the patient and assigned a variable as institutionalized (Yes/No) to differentiate institutionalized and non-institutionalized patients. By doing so, in total of 16472 patients, we noticed that there are 1709 institutionalized patients. Table 3.7 shows the distribution of institutionalized and non-institutionalized patients based on COVID test result (Positive, Negative). Also, for the analysis we are including the SVI and ACS data to only non-institutionalized patients because people who are institutionalized do not reflect the social and demographic characteristics of the neighborhoods

their institution is located in. Moreover, the institutionalized patient population are treated as separate population in the US census data.

Table 3.7. COVID Positive and Negative Counts

Category	Positive	Negative
Institutionalized	554	1,155
Non-Institutionalized	2,204	12,559

3.1.7.3. Variables in COVID Dataset

Table 3.8 shows the list of fields/variables available in the COVID Patient dataset

Table 3.8. Factors in COVID Patient Data

Factors	Variable Type	Category
MRN	String	Patient Data
BMI	Numerical	
Age	Numerical	
Age_Cat	Categorical	
Length of Stay (LOS)	Numerical	
Type	Categorical	
Living_Status	Categorical	
Institutionalized	Categorical	
Institution_name	String	
Institution_group	Categorical	
Week_Year	Categorical	
Sex	Categorical	
Smoking_Status	Categorical	
Race_Ethnicity	Categorical	
Discharge_Disposition	Categorical	
COVID_results	Categorical	
Charlson_Predictive_Mortality	Categorical	
Auto immune	Categorical	
CEVD	Categorical	
CKD	Categorical	
CLD	Categorical	
CVD	Categorical	
Diabetes	Categorical	
ESRD	Categorical	
Hep B	Categorical	
Hep C	Categorical	
HIV	Categorical	
HTN	Categorical	
Immunocompromised	Categorical	
Obesity	Categorical	
Overweight	Categorical	
PVD	Categorical	
Respiratory	Categorical	

Table 3.8. (cont'd) Factors in COVID Patient Data

Factors	Variable Type	Category
RPL_THEME1	Numerical	LA_Census Tract_SVI Data
RPL_THEME2	Numerical	
RPL_THEME3	Numerical	
RPL_THEME4	Numerical	
RPL_THEMES	Numerical	
EP_UNINSUR	Numerical	
EP_GROUPQ	Numerical	
E_TOTPOP	Numerical	
EP_POV	Numerical	
EP_UNEMP	Numerical	
EP_PCI	Numerical	
EP_NOHSDP	Numerical	
EP_AGE65	Numerical	
EP_AGE17	Numerical	
EP_DISABL	Numerical	
EP_SNGPNT	Numerical	
EP_MINRTY	Numerical	
EP_LIMENG	Numerical	
EP_MUNIT	Numerical	
EP_MOBILE	Numerical	
EP_CROWD	Numerical	
EP_NOVEH	Numerical	
Prim_RUCA	Categorical	2010 Census Tract Data
EP_OCC_MBSA	Numerical	American Community Survey
EP_OCC_SER	Numerical	
EP_OCC_SAL_OFF	Numerical	
EP_OCC_NRCM	Numerical	
EP_OCC_PTMM	Numerical	

The COVID positive-tested patient's data are filtered based on the test result ('Positive'). As discussed earlier in this research study, the institutionalized positive patients are treated as separate population. Hence, we have two data sets one is institutionalized COVID positive patient data set and the other dataset is non-institutionalized COVID positive patient data. This split is done based on Institutionalized field ('Yes'/'No'). Appendix G shows the categorical variables and their respective frequency counts of COVID positive institutionalized patients. Appendix H shows the categorical variables and their respective frequency counts of COVID positive institutionalized patients.

From Appendix G data we observe that institutionalized patients who are tested positive for COVID-19 diseases are above age 60 (84%), admitted in hospital (50.7%), coming from Nursing Homes (89.2%), female gender (50.9%), never smoked (48.9%), white or Caucasian (46.4%), Black or African (43.7%), having underlying conditions such as HTN (86.5%), Diabetes (45.3 %),

CVD (36.5%), CKD (30.5%), CEVD (26.7%), Respiratory (20.2%), Obesity (19.1 %), PVD (8.8 %) and, other problems such as Overweight, Immunocompromised, HIV, Hep_B, Hep_C, ESRD, CLD, Auto_immune are less than 3%. This implies the age factor and the underlying conditions such as HTN, Diabetes, CVD are showing more impact on the institutionalized patients to acquire COVID-19. Moreover, 20.25% of FMOLHS COVID institutionalized positive patients are dead.

From Appendix H data we observe that the non-institutionalized patients who are tested positive for COVID-19 diseases are between 21-50 (45%), above age 50 (51.4%), admitted in hospital (28.1%), ED (25.6%), female gender (59.1%), never smoked (66.5%), white or Caucasian (29.6%), Black or African (59.5%), having underlying conditions such as HTN (48.8%), Diabetes (28.3 %), CVD (12.8%), CKD (10.7%), CEVD (6%), Obesity (26.5%), Respiratory (10.2%), and other problems such as PVD, Overweight, Immunocompromised, HIV, Hep_B, Hep_C, ESRD, CLD, Auto_immune are less than 3%. This implies the age factor and the underlying conditions such as HTN, Diabetes, CVD are showing more impact on the non-institutionalized patients to acquire COVID-19. 7.4 % of FMOLHS COVID non-institutionalized positive patients are dead.

3.1.8. Spatial Analysis

In this research study, we analyzed the patterns of number of ED High Utilizers ('Yes/No'), COVID positive tested patients based on Institutionalized ('Yes/No'). For, this we used Spatial Autocorrelation (Global Moran's I) tool. This tool helps us to measure the spatial autocorrelation based on feature locations and the attribute values. The spatial autocorrelation tool calculates Index value, z-score, and p-value to evaluate the significance of index [71].

The Moran's I statistic for spatial autocorrelation is shown by the equation below:

$$I = \frac{n \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{S_0 \sum_{i=1}^{n} z_i^2}$$

where, z_i is the deviation of an attribute for feature i from its mean $(x_i - \bar{X})$, $\omega_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features, and S_0 is the aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j}$$

The z-score for the statistic is computed as*:*

$$Z = \frac{I - E[I]}{\sqrt{V[I]}}$$

where, $E[I]$ *and* $V[I]$ values are:

$$E[I] = \frac{-1}{(n-1)}$$

$$V[I] = E[I^2] - E[I]^2$$

Z-scores represent the number of standard deviations from the mean, and the p-value is a probability that that a sample observation occurs this far from the mean (compared against the chosen test significance level α to determine significance of the test results). If a test's α=0.05, if the observed test result has p-value ≤ 0.05, then the results are statistically significant whereas if

p-value > 0.05, then the results are not statistically significant. In spatial autocorrelation tool, p-value is used to check whether the observed spatial pattern is the result of random process [72]. Moreover, the higher the Z value (its absolute value) or the lower the p value, the more significant the detected spatial autocorrelation pattern is. Also, Moran's Index varies between −1 and +1. A value near +1 indicates that the attributes are clustered; and a value near −1 indicates that the attributes are dispersed. If a Moran's I is close to 0, it indicates a random pattern or absence of spatial autocorrelation [71].

Also, the Hot Spot Analysis is used to identify if high or low values cluster spatially. For this we used the Hot Spot Analysis (Getis-Ord Gi*) tool in ArcGIS software. This tool helps us to identify statistically significant spatial clusters of high values (hot spots) and low values (cold spots). Moreover, it calculates z-score, p-value, and a confidence level bin field (Gi_Bin) [73].

The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j}x_j - \bar{X}\sum_{j=1}^{n} w_{i,j}}{S\sqrt{\frac{n\sum_{j=1}^{n} w_{i,j}^2 - (\sum_{j=1}^{n} w_{i,j})^2}{n-1}}}$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature I and j, n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n}$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\overline{X})^2}$$

The G_i^* statistic is Z-score. Table 3.9 shows different confidence levels for identifying the hot/cold spots.

Table 3.9. Z-score and p-values for Different Confidence Levels

Z-score	p-value	Confidence level	Gi Bin
<-1.65 or >+1.65	<0.10	90%	-1 or 1
<-1.96 or >+1.96	<0.05	95%	-2 or 2
<-2.58 or >+2.58	<0.01	99%	-3 or 3

The positive Z-score and small p-value indicates a spatial clustering of high values (identified as hot spots) and the Gi Bin value will be in the range of 1 to 3. However, negative Z-score indicates spatial clustering of low values (identified as cold spots) and the Gi Bin value will be in the range of -3 to -1. If we cannot reject the null hypothesis then the Gi Bin value is "0" and it is Not Significant. Also, we have another spatial technique called Anselin Local Moran's I Cluster-Outlier analysis and is different from Hot Spot analysis explained earlier. In this analysis, we can identify statistically significant spatial clusters such as High-High (HH) or Low-Low (LL) and outliers such as High-Low (HL) or, Low-High (LH). In this analysis, the high positive Z-score indicates surrounding features have similar values (HH or LL clusters). A low negative Z-score surrounding features have dissimilar values (HL or, LH outliers). If the null hypothesis cannot be rejected then, it is Not Significant. Anselin Local Moran's I Cluster-Outlier analysis is preferred over Hot Spot analysis only if we want to know abnormal trends in an area and to identify spatial outliers. For

example, if we want to identify the features that have high values in cold spot neighborhoods or features that have low values in high spot neighborhoods then Anselin Local Moran's I Cluster-Outlier analysis can be used [74].

3.2. Correlation Analysis and Feature Selection Procedures

SPSS software (version 26) was used to apply statistical techniques in order to analyze the correlation between the dependent and Independent variables mentioned in Table 3.5 and the correlation between the variables in Table 3.8 with respect to COVID result ('Positive/Negative'), Living Status ('Alive/Dead'). In this research study, for studying the relation between numeric dependent variable (ED_Unnecessary_Visits) and other numeric independent variables, Spearman/Pearson correlation techniques can be used.

3.2.1. Pearson Correlation

The Pearson correlation technique is a bivariate analysis useful in studying the strength and direction of the linear relationship between two numeric (continuous) parameters. This can be done by using "Correlate" option in Analyze menu in SPSS. The Pearson correlation coefficient (r) lies between -1 to 1. The negative (-ve) sign indicates that the dependent variables decrease with increase in independent variable and vice versa for positive (+ve) sign. The following assumptions should be satisfied to perform the person correlation.

- All variables should be continuous variables
- Linear relationship between the two variables (using "Chart Builder" option in Graphs in SPSS)
- No significant outliers (using "Explore" option in Analysis in SPSS)
- All the variables should be normally distributed. (This was tested using Shapiro-Wilk test of normality in SPSS).

Table 3.10 shows the Pearson's (r) Correlation Coefficient Scale

Table 3.10. Pearson's Correlation Coefficient Scale [75]

Pearson's Correlation Coefficient	Interpretation
0.00	No association between the two variables
0.01–0.19	No or negligible association between the variables
0.2–0.39	Weak association between the variables
0.4–0.69	Medium association between the variables
0.70–1.0	Strong association between the variables

Though the variables are continuous, there is no linear relationship between the dependent and sample of independent variables. This is shown in the scatterplots shown in Figure 3.3 between dependent (ED_Unnecessary_Year, ED_High Utilizers) and EP_MINRITY independent variable. Moreover, there is no linear relationship between COVID_result ('Positive (1)/'Negative (0)'), Living Status ('Alive/Dead') and EP_MINRITY factor. This is shown in the scatterplots shown in Figure 3.4 . Hence, the Pearson correlation score might not be appropriate with this analysis.

3.2.2. Spearman Correlation

The Spearman rank-order correlation (Spearman correlation r_s) was used to find the correlation between two ordinal or two continuous variables. Spearman is a nonparametric measure whereas Pearson is a parametric measure. The range of r_s is same as r discussed in Section 3.2.1. This can be done by using the "Correlate" option in Analyze menu in SPSS. This analysis is used when the data violates the assumptions of Pearson. The following assumptions should be satisfied to perform the Spearman correlation analysis.

- The variables should be ordinal or scale.
- The variables can be monotonically related to another variable. This implies if one variable increases the other variable should either increase or decrease (using "Chart Builder" option in Graphs in SPSS)

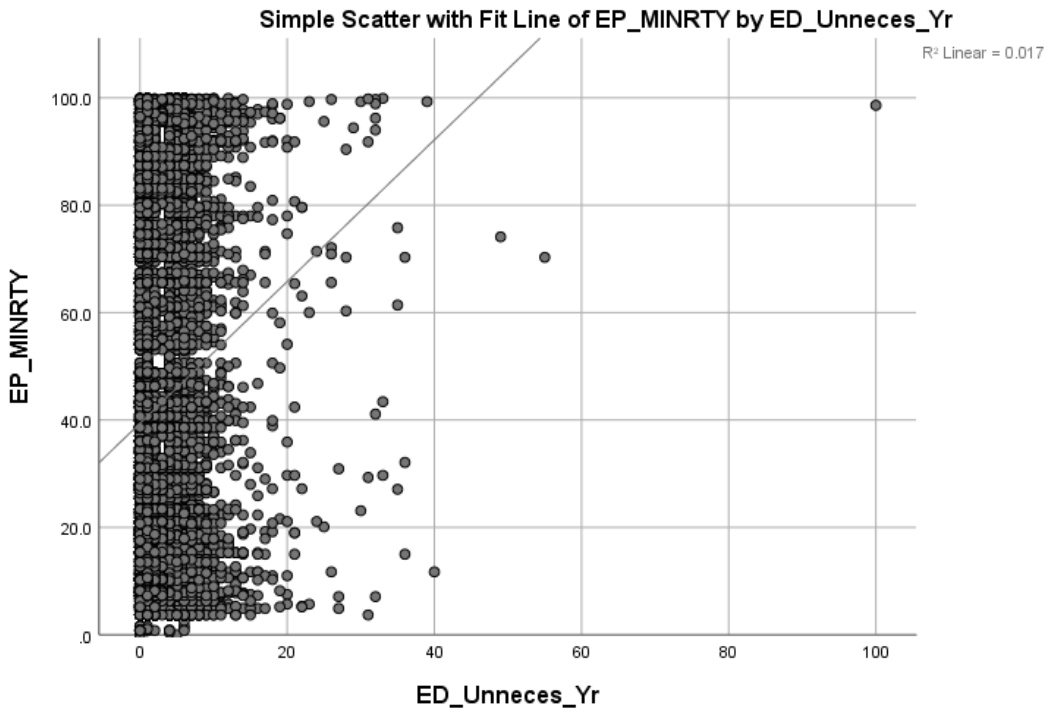
The variables are scale variables and some of the independent variables either increased or decreased with the increase in dependent variable. This is shown in the scatterplots shown in Figure 3.3 and Figure 3.4. Hence, we can use the Spearman correlation analysis.

3.2.3. Eta Coefficient Test

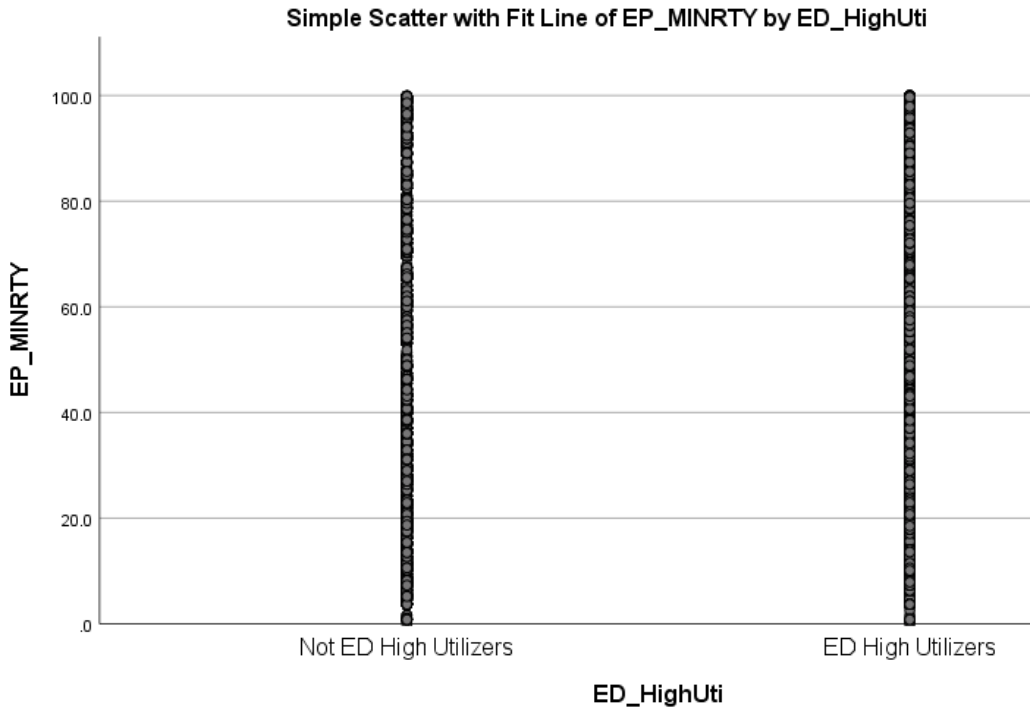
The Eta Coefficient test is a statistical method used to find the association between an independent categorical variable (e.g., gender, race), and a dependent scale variable (e.g., a continuous value such as 0, 1, 2, and so on) [75]. The Eta Coefficient is asymmetric and is used to find the nonlinear association between a categorical and a scale variable. Moreover, the association score criteria are same as Pearson's correlation coefficient discussed in Table 3.10. The following are assumptions of the Eta Coefficient test [75].

- The data must be nonlinear or curvilinear variables (using "Chart Builder" option in Graphs in SPSS)
- The data must be asymmetric (Using "Descriptive Statistics->Frequencies" option in Analyze in SPSS)
- The dependent variable should be scale or interval level
- The independent variables should be categorical with two or more
- There must be independence of observations, so there is no relationship between the groups or between the observations in each group [75].

In Figure 3.3 and Figure 3.4, we can see that the sample of data is nonlinear. The data in the ED visits dataset and COVID dataset is asymmetric. This is shown in the histograms shown in Figure 3.5 and Figure 3.6 for sample of the factors.

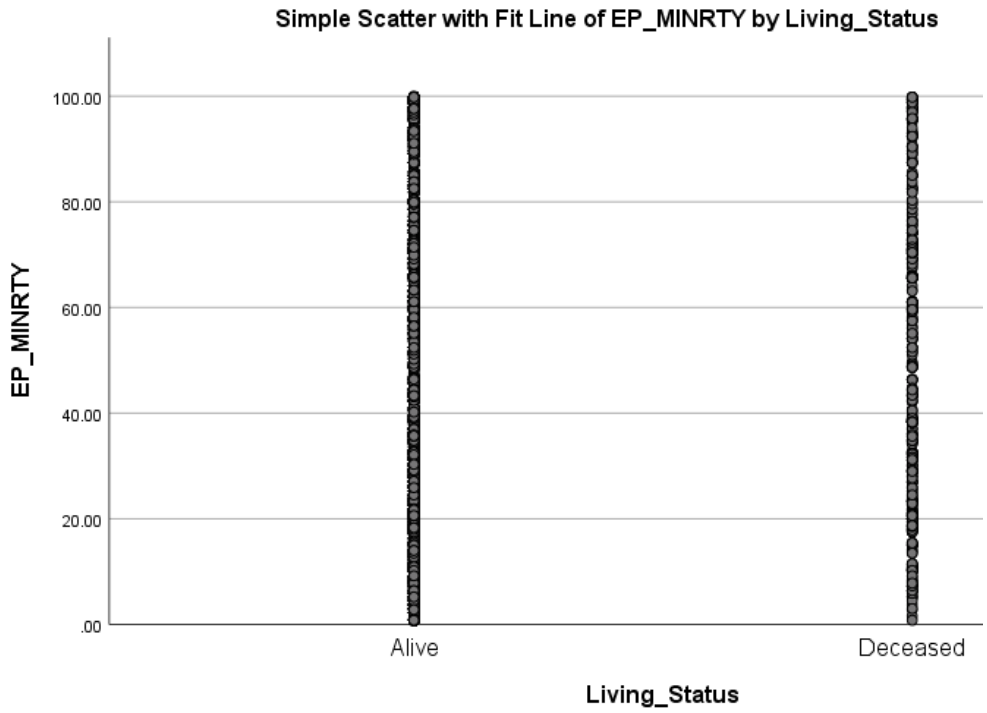


EP_MINRTY vs ED_Unnecessary_Yr

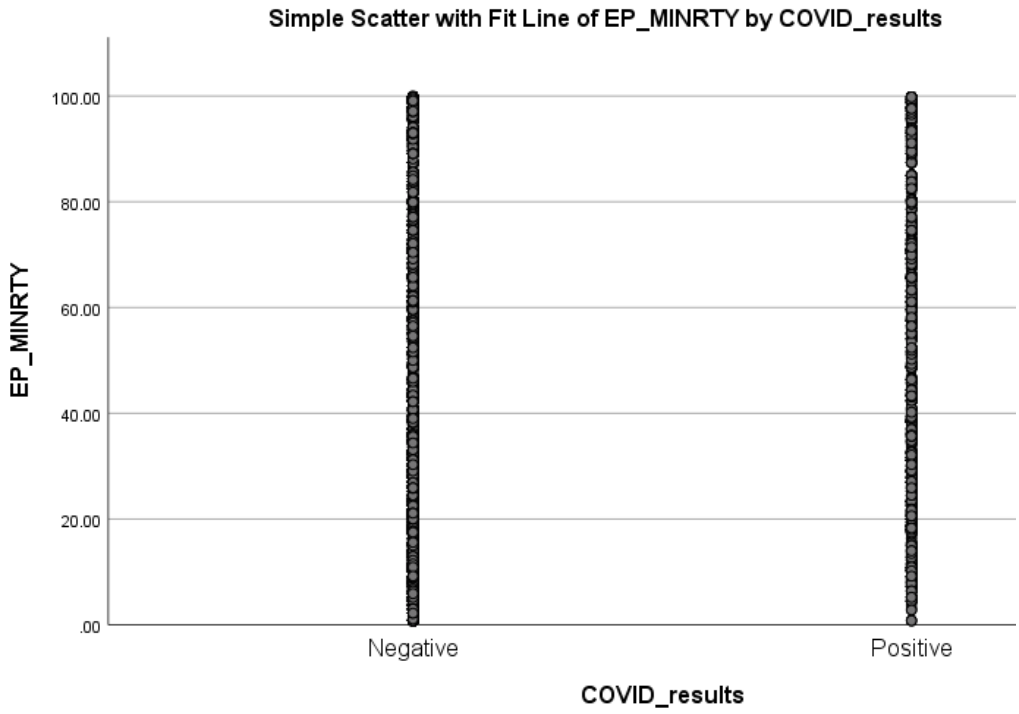


EP_MINRTY vs ED_HighUtilizers

Figure 3.3. ED data scatter plots between Dependent and Independent Variables



a) EP_MINRTY vs Living_Status



b) EP_MINRTY vs COVID_results

Figure 3.4. COVID data scatter plots between Dependent and Independent Variables

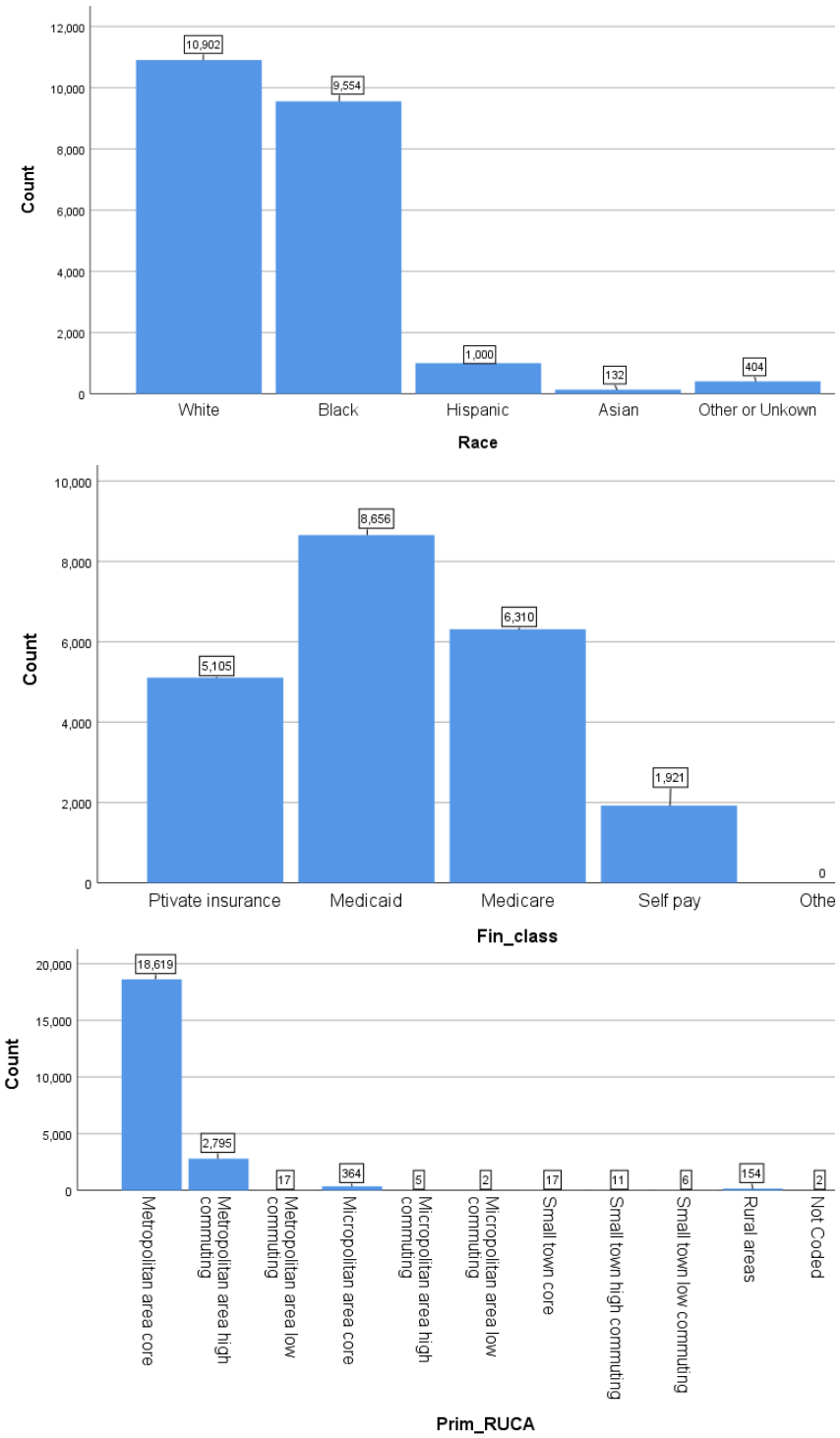


Figure 3.5. ED Visit Data Histograms

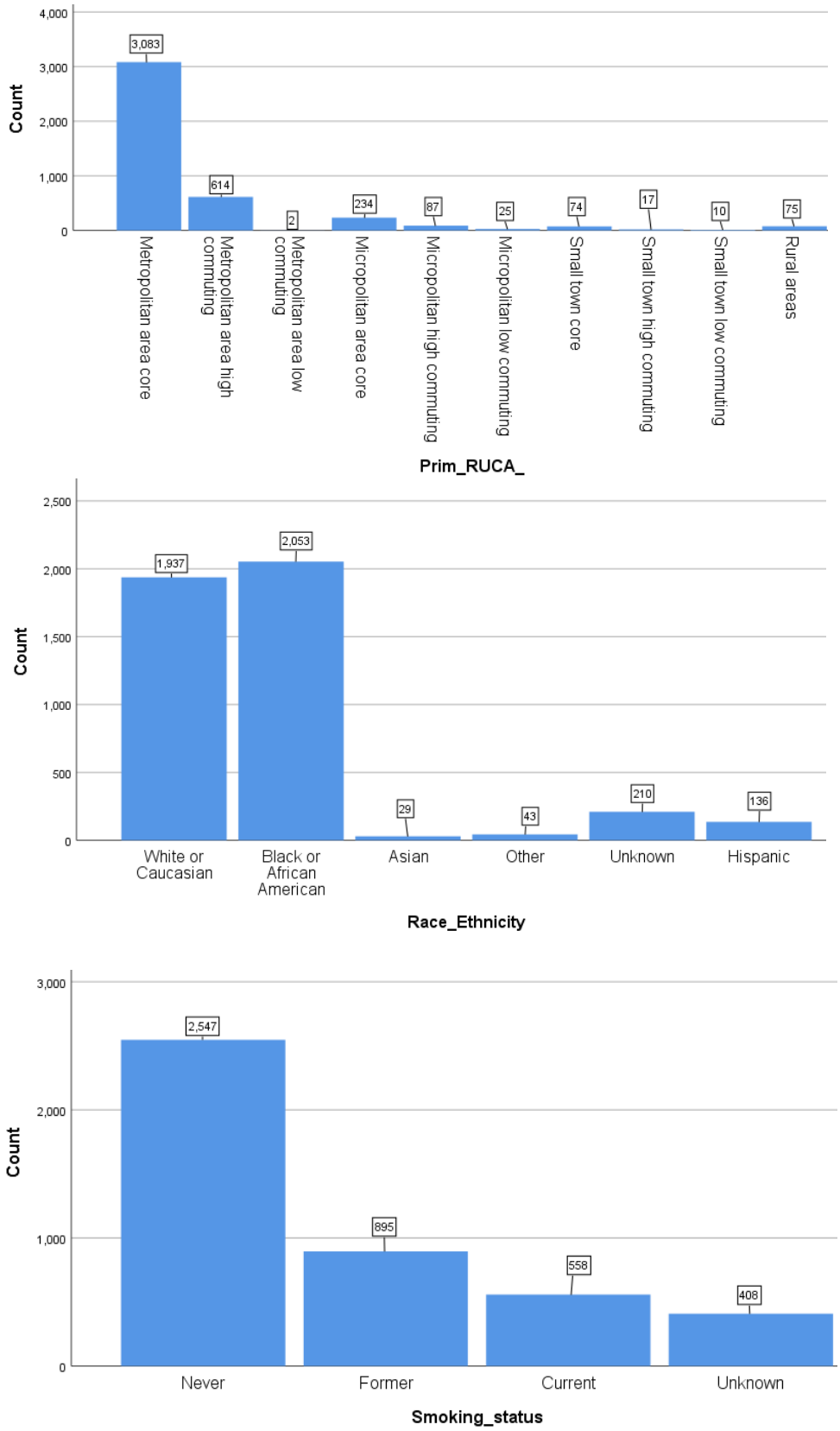


Figure 3.6. COVID Data Histograms

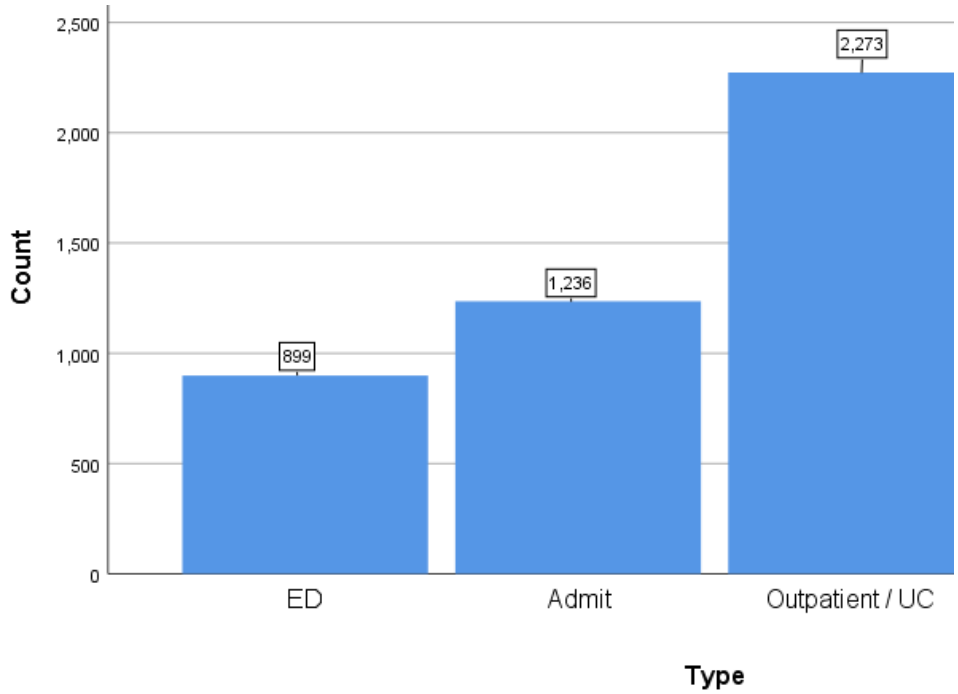


Figure 3.6 (cont'd) COVID Data Histograms

3.3. ED Predictive Model Development

Van Kuijk et al [76] indicates that predictive modelling is used in developing tools for prediction of the most likely value of a continuous measure or the probability of the occurrence of an event; predictive methods include classification, regression, and time series modeling. This research study considers that a patient from each census tract/block group may visit (once, or more than once) different ED in LA for various reasons. The output of the model is to predict whether a patient is likely to be an ED high utilizer ('Yes/No') of ED facilities affiliated to FMOLHS that are available in LA. The research question addressed here is:

- Develop and validate a machine learning / neural network model for predictive classification of ED high utilizers as (Yes/No) to multiple ED facilities available in LA and identify correlations between the SDOH factors and the other independent variables influencing the utilization of FMOLHS ED.

In this research study, we are considering the FMOLHS unique ED patients between January 2015- January 2020 from all the cities in Louisiana. The geocoded patient data is imported into ArcGIS software as point layer data using the latitude and longitude coordinates of each patient's address. Also, the SVI data shapefile is imported into the ArcGIS software. By using spatial join query between the point layer of patient visits dataset and SVI polygon layer, each patient is mapped with the SVI data related factors based on the patient address. Now, the entire patient data mapped to the SVI data is exported from the attribute table. For this research objective, we are using the unique ED High Utilizers data.

Before developing the model, to understand the influence of independent variables on the dependent variables shown in Table 3.5, correlation analysis is performed using Spearman correlation and Eta Coefficient test statistical techniques. As discussed in Section 3.1.6.1, the

unique ED High Utilizers (10,996) patient records obtained from FMOLHS affiliated hospitals in LA between 2015-2019 are added to the randomly selected 10,996 patient records who are not ED High Utilizers. This dataset is used to obtain the correlation scores. Moreover, feature selection of independent variables is done by considering the correlation scores. In this scenario, the independent variables which has correlation score as 0 with respect to the dependent variable are unrelated and removed. Independent variables which have correlation score other than 0 with respect to the dependent variable were considered for analyzing the feature importance among the independent variables. Also, if there were multiple independent variables highly correlated with each other, then we selected only one variable among them and eliminated the rest of the variables. Finally, ranks were assigned to the independent variables based on the correlation score with respect to dependent variable; the results are discussed in Section 4.1. For example, rank 1 is assigned to the variable that has high correlation score and the least ranking variables are removed from the dataset.

The machine learning task is a classification task. To predict the ED_High Utilizers dependent variable we considered the unique "ED Discharge" patient visits filtered based on MRN. In the process of developing the ED High Utilizers predictive classification model, we have randomly selected 10,996 patient records who are not ED High Utilizers and added to the ED High Utilizers dataset to make the model learn the difference between ED High Utilizers and Non-High Utilizers. The correlated independent variables (factors) identified are added to the two prediction models.

3.3.1. Preprocessing

To perform machine learning-specific preprocessing and develop the predictive classification models, Python version 3.7 was used. We imported the FMOLHS ED patient visits dataset with the help of the Pandas Python library.

The StandardScaler method from the Python sklearn library, sklearn.preprocessing, was used to normalize input variable values to have a mean of 0 and standard deviation of 1.

SPSS was used to perform dummy coding of the categorical variables. By clicking on the "Variable View" tab and navigating to "Values" column the categorical variables are manually converted to numerical codes such as 0,1,2. Additionally, the nulls/Not Applicable values are categorized as" Missing data/Unknown/Other" and converted to numerical codes. The codes assigned to the values in the categorical variables are shown in Appendix F.

The Python codes for the three machine learning models including preprocessing and validation are shown in Appendix J, Appendix K, and Appendix L.

3.3.2. Neural Networks

Neural networks represent sets of neurons which are connected. Generally, neurons are a non-linear parameterized function of its input variables and its value is its output. Therefore, a network of neurons is the composition of the nonlinear functions of two or more neurons [77]. The simple neural network structure is shown in Figure 3.7. From Figure 3.7 we can see that neuron like structure is presented with an input layer, hidden layer, and an output layer. The input vector goes through weighted connections, intermediate nodes, and then the weights are learned in order to obtain correct outputs. The neural network machine learning algorithm is capable of learning complex relationships with the help of weighted connections.

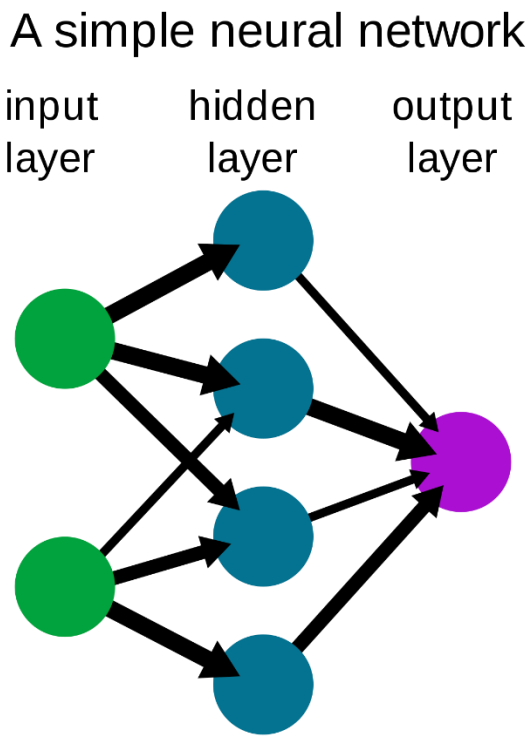


Figure 3.7. Neural Network; Source Image: Neural network example.svg, Wikipedia [78]

We imported MLPClassifier from sklearn.neural_network to train the model to predict the dependent variable (ED_High Utilizers). In this scenario we defined the following parameters for the MLPClassifier [79]

- Input layer: The number of features in the dataset (20) and bias (1) are considered as number of nodes in input layer (number of nodes:21)
- Hidden layer: hidden layer sizes (number of layers:3, number of nodes in each layer: 20)
- Output layer: The two outcomes of the target variable are considered as number of nodes in output layer (number of nodes:2; one node corresponds to ED-High Utilizer (1) and the other one corresponds to Not ED High Utilizer (0))
- Activation function for the hidden layers is the rectified linear unit function, returns f(x) = max (0, x) (activation: 'relu')
- MLPClassifier supports multi-class classification by applying 'Softmax' (It is a generalization of the logistic function to multiple dimensions) as the activation function for output layer.
- The stochastic gradient descent solver is used for weight optimization (solver: 'sgd)'. This solver calculates the error derivative (contribution) of each weight and for each training data instance and calculate the update of weight values to reduce future errors.
- Regularization term on weights (alpha=0.3). This reduces the overfitting by constraining the magnitude of weights.
- The learning rate for weight updates (learning_rate='adaptive'). keeps the learning rate changing. The limit to the weight changes is 'learning_rate_init'.

- The initial learning rate used (learning_rate_init=0.3). This limit is used to control the amount of weight change during training of the model.
- Momentum for gradient descent update. The momentum is used to avoid or move out of the local minima. This will eventually reduce the error function to reach the global minima. (momentum=0.4).
- The maximum number of iterations (max_iter=1000); for solver=’sgd’ it indicates the maximum number of times the training dataset is repeatedly applied. Learning may stop earlier if convergence is reached.

3.3.3. Random Forests

Random forest is a supervised learning technique and uses decision tree model for parametrization, but it integrates a sampling technique, a subspace method, and an ensemble approach to optimize the model building. It can be applied to both classification and regression problems [80]. Moreover, the random forest algorithm is based on bagging approach. Bagging is a way to decrease the variance in the prediction by generating additional data for training from dataset using combinations with repetitions to produce multi-sets of the original data. This algorithm is bootstrapping the data by randomly choosing subsamples for each iteration of growing trees. The growing happens in parallel which is a key difference between XGBoost and random forests [81,82,83]. The characteristics of random forest algorithm are shown in Figure 3.8.

Figure 3.8. Characteristics of Random Forest

We imported the RandomForestClassifier from sklearn.ensemble to train the model to predict the dependent variables. In this scenario we defined the following parameters for the RandomForestClassifier [84].

- The number of trees in the forest (n_estimators=1800)
- The function to measure the quality of split. Here we used “entropy” for the information gain (criterion='entropy')

- The maximum depth to which the nodes are expanded (max_depth=90)
- The minimum number of samples required to split an internal node in the decision tree (min_samples_split=2)
- The minimum number of samples required at leaf nodes in the decision tree (min_samples_leaf=1)
- The number of features considered when looking at best split (max_features='auto'); for auto max_features=sqrt(n_features)
- Bootstrap sampling (smaller sample that is bootstrapped from larger sample) is used for building a tree (bootstrap='True')
- This random_state field value is used to control both the randomness of the bootstrapping of the samples used when building trees (if bootstrap=True) and the sampling of the features to consider when looking for the best split at each node (if max_features < n_features). (random_state=12)

3.3.4. XGBoost

In XGBoost individual decision trees are created using multiple cores to improve training time performance. Decision trees are also organized in a way that reduces lookup times when the trained model is applied. XGBoost increases the performance and decreases the training time of the models [85]. This is a relatively new algorithm and is utilizing the concept of gradient tree boosting and use trees in a sequential learning process as weak learners. Boosting is an iterative technique which adjusts the weight of an observation based on the last classification [81,82]. Moreover, by introducing regularization parameters we can reduce overfitting. The characteristics of XGBoost algorithm are shown in Figure 3.9.

Figure 3.9 Characteristics of XGBoost

We imported XGBClassifier from XGBoost in order to train the model to predict the dependent variables. In this scenario we defined the following parameters for the XGBClassifier [86].

- Global bias, the initial prediction score of all instance (base_score=0.5)
- Subsample ratio of the training instances (subsample=1)
- Colsample_bylevel is the subsample ratio of columns for each level. Subsampling occurs once for every new depth level reached in a tree. Columns are subsampled from the set of columns chosen for the current tree. (colsample_bylevel=1)
- Colsample_bynode is the s the subsample ratio of columns for each node (split). Subsampling occurs once every time a new split is evaluated. Columns are subsampled from the set of columns chosen for the current level. (colsample_bynode=1)
- Colsample_bytree is the subsample ratio of columns when constructing each tree. Subsampling occurs once for every tree constructed. (colsample_bytree=0.9)
- Minimum loss reduction required to make a further partition on a leaf node of the tree. If the gamma value increases the algorithm becomes more conservative (gamma=0.1)
- To know the contribution of each feature for each tree in the model (importance_type='gain')
- The weight update used to prevent overfitting: range [0,1] (learning_rate=0.05)
- Maximum delta step we allow each leaf output (max_delta_step=0)
- Maximum depth of a tree (max_depth=12)
- Minimum sum of instance weight needed in a child (min_child_weight=5)
- The number of trees build (n_estimators=100)
- Number of parallel trees constructed during each iteration. This option is used to support boosted random forest. (num_parallel_tree=1)
- The learning task and corresponding learning objective (objective='binary: logistic')
- Control the balance of positive and negative weights, useful for unbalanced classes (scale_pos_weight=1)

3.3.5. Training, Testing, and Validation

The dataset was randomly divided into a training and testing set and a validation set. The training and testing set together contains 70% (15,394 records) of data and validation set contains 30% (6,598 records) of data. Using SPSS->Data->Select Cases->Random Sample of Cases we have randomly selected 30% of total unique ED High Utilizers (10,996) and selected output as “Filter out unselected cases”. In SPSS with the help of filter column we separated the filtered data as validation dataset (30%) and remaining data as a training and testing dataset (70%). Likewise, from the 408,432 Not-ED High Utilizers we have randomly selected 10,996 Not-ED High Utilizers and separated the validation and a training and testing dataset. Finally, the corresponding ED High Utilizers validation and a training and testing data is merged with Not ED High Utilizers validation and a training and testing data. Hence, the training and testing dataset and the validation dataset have equal number of ED High Utilizers (1) and Not-ED High Utilizers (0). The validation dataset is used to validate the machine learning techniques utilized; the same data should not be used in both training and validation of the model. During training, KFold (from sklearn.model_selection; the value of K is 10) is used to break the training and testing dataset randomly and repeatedly into training and testing datasets, a technique that help improve generalization and reduce overfitting. The developed models are trained on the training dataset and the evaluation metrics Area Under

Curve (AUC) score, Accuracy, Precision, Recall, Specificity and F1 score are used to evaluate the models. The models which have high AUC and F1 score are the best fit models. Additionally, the models should have higher Accuracy, Recall, Specificity, Precision scores as they are used to examine the percentage of number of correctly classified ED high utilizers, Not-ED high utilizers with respect to actual count of ED high utilizers, Not-ED high utilizers in the dataset. Table 2.1 in Chapter 2 shows the calculation details for each evaluation metric.

3.4. ED High Utilizers Spatial Analysis

After the predictive classification model development, the ED high utilizers data was used to perform spatial analysis. Spatial Autocorrelation, Hot Spot Analysis as discussed in Section 3.1.8 was performed to identify the clusters of ED High Utilizers based on unique FMOLHS ED patient data between 2015-2019.

3.5. Retrospective Study of COVID Tested Patients

Due to the COVID outbreak in the United States and increase in the COVID cases in LA, FMOLHS asked for a retrospective study to identify factors which may be responsible for increased risk of FMOLHS patients acquiring COVID-19 disease. The research question addressed here is:

- Perform a retrospective study and spatial analysis to identify which patient demographics, clinical characteristics and SDOH factors are responsible for increasing the risk of general population acquiring or dying due to COVID-19.

In this research study, to perform of the retrospective study of COVID tested patients is separated into four datasets. The first filter applied is based on the on the test result ('Positive/Negative') for SARS-CoV-2. The First dataset contains all positive tested patient's data, and the other dataset contains all negative tested patients. As discussed in Section 3.1.7.3, the positive tested patient's data is separated into another two datasets based on Institutionalized field ('Yes'/'No'). The third data set contains all positive institutionalized patient's data and the fourth one contains all positive non-institutionalized patients. All dataset records were geocoded into latitude and longitude coordinates based on the institution's address for institutionalized patients and patient's physical address for non-institutionalized patient's. The geocoded datasets are imported into ArcGIS software as point layer. The SVI and ACS data are imported into the ArcGIS software as shapefile. The geocoded patient data was then related to the Census data (SVI and ACS) discussed in Section 3.1.3 with the help of spatial joining technique available in the ArcGIS tool. The research question is addressed by comparing the datasets as shown below:

- Correlation analysis was performed between positive and negative tested patients using Spearman Correlation and Eta Coefficient test statistical techniques and the features that have high influence on the COVID test result, Living Status are identified based on the feature selection process discussed in Section 3.3.
- To study the differences and similarities between positive and negative tested patients for SARS-CoV-2, the variables in the dataset are summarized (averages, counts) by Age_Category, Race_Ethnicity, Sex, Living_Status, Smoking_Status, Type, Institutionalized, Institution_Group. Also, Hot Spot Analysis as discussed in Section

3.1.8 was performed to identify the clusters of positive and negative FMOLHS COVID tested patients in LA.

- To study the differences and similarities between positive institutionalized and non-institutionalized patients, the variables in the dataset are summarized (averages, counts) by Age_Category, Race_Ethnicity, Sex, Living_Status, Smoking_Status, Type, Institutionalized, Institution_Group. Also, Spatial Autocorrelation, Hot Spot Analysis as discussed in Section 3.1.8 was performed to identify the clusters of positive institutionalized and non-institutionalized FMOLHS COVID tested patients in LA.

Chapter 4. ED Prediction Model Results and Analysis

4.1. Correlation Analysis and Feature Selection

To study the association between the independent and dependent variables discussed in Section 3.1.6, the Spearman correlation and Eta Coefficient test statistical techniques are used. For this we considered all the unique ED High Utilizers patient data (10,996 records) between 2015-2019 years along with another randomly selected ED visit patient data that are not high utilizers. This data has been selected randomly with the help of SPSS Data->Select Cases->Random sample of cases. The Spearman correlation is performed between *numeric(continuous/binary/Ordinal)* dependent and independent variables as the dataset satisfied the test assumptions. The H_0 (null hypothesis) is defined as "There is no[monotonic] association between the variables".
The following SPSS code is used to run the Spearman correlation analysis.

```
NONPAR CORR
  /VARIABLES=Year LOS_Yr Age Admit Gender UC_Visits PC_Visits
ED_Unneces_Yr ED_HighUti E_TOTPOP
    EP_POV EP_UNEMP EP_PCI EP_NOHSDP EP_AGE65 EP_AGE17
EP_DISABL EP_SNGPNT EP_MINRTY EP_LIMENG EP_MUNIT
    EP_MOBILE EP_CROWD EP_NOVEH EP_GROUPQ RPL_THEME1
RPL_THEME2 RPL_THEME3 RPL_THEME4 RPL_THEMES
    EP_UNINSUR
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

The results of the Spearman Correlation analysis are shown in Table 4.1.
From Table 4.1, there is a strong correlation between the two dependent variables, ED_Unnecessary_Yr and ED_High_Utilizers. In this scenario, we have selected the ED High Utilizers dependent variable over ED_Unnecessary_Visits because the ED_High Utilizers has showed higher correlation scores with respect to the independent variables when compared to ED Unnecessary Visits.

Considering the correlation scores between dependent (ED_High Utilizers) variable and independent variables, the E_TOTPOP, EP_AGE17, EP_LIMENG, EP_MOBILE, RPL_THEME3 independent variables showed negligible correlation scores with respect to dependent variables. So, removing these variables.

From the covariance matrices shown in Appendix I we find that some of the independent variables are highly correlated (correlation scores between 0.7-1) with each other:

- Admit is strongly correlated with LOS_Yr.
- EP_PCI is strongly correlated with EP_POV, EP_NOHSDP
- EP_NOVEH is strongly correlated with EP_POV, RPL_THEME1, RPL_THEMES
- EP_POV, EP_UNEMP, EP_PCI, EP_NOHSDP, EP_NOVEH, RPL_THEMES variables are strongly correlated with RPL_THEME1.
- EP_MINRTY, EP_LIMENG are strongly correlated with RPL_THEME3
- EP_SNGPT, EP_DISABL is strongly correlated with RPL_THEME2
- RPL_THEME4, EP_UNINSUR is strongly correlated with RPL_THEMES

We removed all RPL_THEME* variables because they are strongly correlated with other independent variables. EP_PCI is removed as it is strongly correlated with EP_POV,

EP_NOHSDP. EP_POV is selected over EP_NOVEH and Admit is considered over LOS_Yr because the EP_POV, Admit has desirable high correlation scores with the dependent variables when compared to EP_NOVEH, LOS_Yr respectively.

Table 4.1. Spearman Correlation Scores Between Numeric Variables

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	ED_Unnecessary_Yr		ED_High_Utilizers	
	r_s	p-value	r_s	p-value
Year	-0.38	0.000	-0.40	0.000
LOS_Yr	-0.68	0.000	-0.47	0.000
Age	-0.34	0.000	-0.34	0.000
Admit	-0.68	0.000	-0.47	0.000
Gender	-0.01	0.305	-0.01	0.225
UC_Visits	0.15	0.000	0.16	0.000
PC_Visits	-0.12	0.000	-0.14	0.000
ED_Unneces_Yr	***		0.89	0.000
ED_HighUtilizers	0.89	0.000	***	
E_TOTPOP	-0.07	0.000	-0.09	0.000
EP_POV	0.23	0.000	0.25	0.000
EP_UNEMP	0.21	0.000	0.22	0.000
EP_PCI	-0.27	0.000	-0.30	0.000
EP_NOHSDP	0.21	0.000	0.22	0.000
EP_AGE65	-0.17	0.000	-0.17	0.000
EP_AGE17	0.06	0.000	0.05	0.000
EP_DISABL	0.12	0.000	0.13	0.000
EP_SNGPNT	0.16	0.000	0.18	0.000
EP_MINRTY	0.10	0.000	0.12	0.000
EP_LIMENG	-0.05	0.000	-0.05	0.000
EP_MUNIT	0.10	0.000	0.12	0.000
EP_MOBILE	-0.03	0.000	-0.06	0.000
EP_CROWD	0.16	0.000	0.18	0.000
EP_NOVEH	0.17	0.000	0.18	0.000
EP_GROUPQ	0.12	0.000	0.13	0.000
RPL_THEME1	0.25	0.000	0.26	0.000
RPL_THEME2	0.10	0.000	0.11	0.000
RPL_THEME3	0.03	0.000	0.04	0.000
RPL_THEME4	0.22	0.000	0.23	0.000
RPL_THEMES	0.23	0.000	0.25	0.000
EP_UNINSUR	0.27	0.000	0.29	0.000

To find the correlation scores between *numeric* (continuous/binary) dependent and *categorical* independent variables, an Eta Coefficient test is run. The *null hypothesis* is defined as H_0 : There is no correlation between two variables. The results are shown in Table 4.2.

The following SPSS Code is used to find the association between the dependent and independent variables using Eta Coefficient test.

```
CROSSTABS
  /TABLES=Race Prim_RUCA Fin_class BY ED_Unneces_Yr ED_HighUti
  /FORMAT=AVALUE TABLES
  /STATISTICS=ETA
  /CELLS=COUNT
  /COUNT ROUND CELL
  /BARCHART.
```

Table 4.2. Results of the Eta Coefficient Test

Eta ($\dot{\eta}$)- Nominal by Interval Measure		
Independent/Dependent	ED_Unnecessary_Visits	ED_High_Utilizers
Fin_class	0.38	0.51
Race	0.11	0.13
Prim_RUCA	0.05	0.07

From Table 4.2, ED High Utilizers shows higher association scores with respect to the categorical independent variables when compared to ED_Unnecessary_Visits.

- There is a strong association between ED_High_Utilizers and Fin_Class, the Eta coefficient $\dot{\eta} = 0.51$, $\dot{\eta}^2 = 0.26$, implies that we can reject the null hypothesis as there an association between Fin_Class and ED_High_Utilizers; Financial Class has 26% effect on the ED_High_Utilizers.
- There is a weak association between Race and ED_High_Utilizers, the Eta coefficient $\dot{\eta} = 0.11$, $\dot{\eta}^2 = 0.0121$, this implies that we can reject null hypothesis as there an association between Race and ED_High_Utilizers; Race has 1.21% effect on the ED_High_Utilizers.
- There is a weak association between Prim_RUCA and ED_High_Utilizers, the Eta coefficient $\dot{\eta} = 0.07$, $\dot{\eta}^2 = 0.0049$, this implies that we can reject null hypothesis as there an association between Prim_RUCA and ED_High_Utilizers; Prim_RUCA has 0.49% effect on the ED_High_Utilizers.

Based on the Spearman and Eta Coefficient test association scores, feature selection of the independent variables is performed. Rank 1 is assigned to the variable that has highest correlation score among the independent variables and the rank increases if the correlation score decreases. As discussed earlier, we are removing the strongly correlated variables RPL_THEME1, RPL_THEME2, RPL_THEME3, RPL_THEME4, RPL_THEMES EP_PCI, EP_NOVEH, LOS_Yr. In this research study, due to consolidation of multiple visits of a specific patient to a single record and variability in the dataset we obtained lower correlation scores for statically significant variables and decided to accept the variables that showed correlation score greater than 0.1 (irrespective of +/-). We removed independent variables that had correlation scores of less than 0.1 (irrespective of + or – sign, because it only indicates direction); these were E_TOTPOP, EP_AGE17, EP_LIMENG, and EP_MOBILE. The feature ranking of the remaining independent variables with respect to ED High Utilizers dependent variable is shown in Table 4.3.

Table 4.3. Feature Ranking of Independent Variables

Independent/Dependent	ED_High_Utilizers	Rank
Fin_Class	0.513	1
Admit	-0.472	2
Year	-0.400	3
Age	-0.335	4
EP_UNINSUR	0.293	5
EP_POV	0.254	6
EP_UNEMP	0.223	7
EP_NOHSDP	0.216	8
EP_CROWD	0.176	9
EP_SNGPNT	0.175	10
EP_AGE65	-0.173	11
UC_Visits	0.159	12
PC_Visits	-0.144	13
Race	0.134	14
EP_GROUPQ	0.131	15
EP_DISABL	0.125	16
EP_MINRTY	0.118	17
EP_MUNIT	0.115	18
Prim_RUCA	0.066	19
Gender	-0.008	20

Prim_RUCA and Gender fields had negligible correlation scores and were removed.

4.2. ED Prediction Model Development and validation

As discussed in Section 3.3, by using Python 3.7 software the predictive classification models are developed by applying Neural Networks, Random Forest and XGBoost machine learning techniques. The dependent variable/predictor variable is ED_High Utilizers ('Yes (1)/No (0)') and categorical (or binary). The machine learning task is a classification task. Initially the dataset is divided into a training and testing (70%) and a validation (30%) sets. The training and testing (70%) dataset is then divided randomly and repeatedly into training and testing datasets by using K-fold cross validation technique and the value of K is 10. The dependent variable is predicted by including all the independent variables selected as part of feature selection process. The codes for each machine learning methodology are shown in Appendix J, Appendix K, and Appendix L. The average accuracy score, F1-Score, and ROC_AUC for the 10 folds was calculated for each of the different machine learning models and is shown in Table 4.4. The validation dataset is used and the evaluation metrics Accuracy, Sensitivity (Recall), Specificity, Precision, F1-Score and ROC_AUC Score are calculated for the developed models. With the help of these evaluation metrics the models are evaluated. The evaluation metrics are shown in Table 4.5.

In Table 4.5, we can see that the prediction model developed using XGBoost machine learning technique showed slightly higher scores in all the evaluation metrics when compared to other models.

The FMOLHS ED Visit population has large amount of Not ED High Utilizers when compared to ED High Utilizers. We also validated the developed machine learning models on the unbalanced general FMOLHS ED population, and the results are shown in Table 4.6. In the general population validation data set the Neural Network machine learning technique showed slightly higher scores in all the evaluation metrics when compared to other models.

Table 4.4. Average of Accuracy Scores for 10-Folds on Testing Dataset

Classification Technique	Accuracy (%)	F1-Score (%)	ROC_AUC (%)
Neural Network	87.75	87.66	95.13
XGBoost	88.01	87.93	95.29
Random Forest	86.33	86.30	93.96

Table 4.5. Evaluation Metrics for the Prediction Models on Validation Dataset

Classification Technique	Accuracy (%)	Recall	Precision	Specificity	F1 Score (%)	ROC_AUC (%)	Confusion Matrix $\begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}$
Neural Network	87.45	0.87	0.88	0.88	87.33	95.01	$\begin{bmatrix} 2901 & 398 \\ 453 & 2846 \end{bmatrix}$
XGBoost	87.47	0.87	0.88	0.88	87.36	95.11	$\begin{bmatrix} 2913 & 386 \\ 441 & 2858 \end{bmatrix}$
Random Forest	86.42	0.86	0.87	0.87	86.32	94.02	$\begin{bmatrix} 2874 & 425 \\ 471 & 2828 \end{bmatrix}$

Table 4.6. Evaluation Metrics for the Prediction Models on General FMOLHS ED Population

Classification Technique	Accuracy (%)	Recall	Precision	Specificity	F1 Score (%)	ROC_AUC (%)	Confusion Matrix $\begin{bmatrix} TN & FP \\ FN & TP \end{bmatrix}$
Neural Network	50.78	0.86	0.01	0.50	2.79	73.89	$\begin{bmatrix} 202305 & 198400 \\ 446 & 2853 \end{bmatrix}$
XGBoost	50.25	0.87	0.01	0.50	2.77	72.02	$\begin{bmatrix} 200191 & 200514 \\ 441 & 2858 \end{bmatrix}$
Random Forest	49.70	0.86	0.01	0.49	2.71	72.65	$\begin{bmatrix} 197980 & 202725 \\ 471 & 2828 \end{bmatrix}$

4.3. Spatial AutoCorrelation

As discussed in Section 3.1.8, to understand the patterns of unique ED High Utilizers (10996) between 2015-2019 in LA we used Spatial Statistics Tools->Analyzing Patterns-> Spatial Autocorrelation (Moran's I) in ArcGIS. To run this analysis, we have included the census tracts that have positive patients count greater than 0.

- We chose EUCLIDEAN distance method as the study area is small (Example: Census Tracts) and it calculates the distance between two points connected by a straight-line.

- We chose conceptualization of spatial relationships as INVERSE DISTANCE because the nearby features have a larger influence on the computations of a target feature than features that are far away.

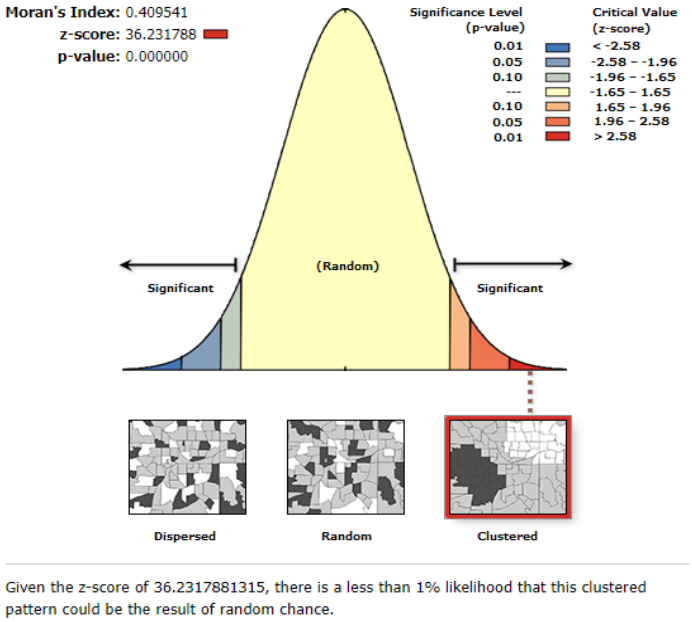
The metric reported are shown in Figure 4.1, the values are shown in Table 4.7

Table 4.7. Global Moran's I Results

Global Moran's I Summary	ED High Utilizers
Moran's Index:	0.409541
Expected Index:	-0.003067
Variance:	0.000130
Z-score:	36.231788
p-value	0.000000

From Table 4.7 we can see that

- The null hypothesis in this analysis is defined as "the values associated with features are randomly distributed". Here P-value is near 0 (<0.01) and statistically significant for ED High Utilizers. So, null hypothesis can be rejected.
- Moran's Index varies between −1 and +1. A value near +1 indicates that similar attributes are clustered (either high values near high values or low values near low values); and a value near −1 indicates that dissimilar attributes are clustered (either high values near low values or low values near high values). If a Moran's I is close to 0, it indicates a random pattern or absence of spatial autocorrelation.
    - **ED High Utilizers**: The Z-score for Moran's I is 36.231788 and p-value is 0.00, these values indicate that the detected spatial autocorrelation (ED High Utilizers) was statistically significant. In this scenario, Moran's I (0.409541) is close to +1, it indicates that the ED High Utilizers census tracts in the study area are clustered.

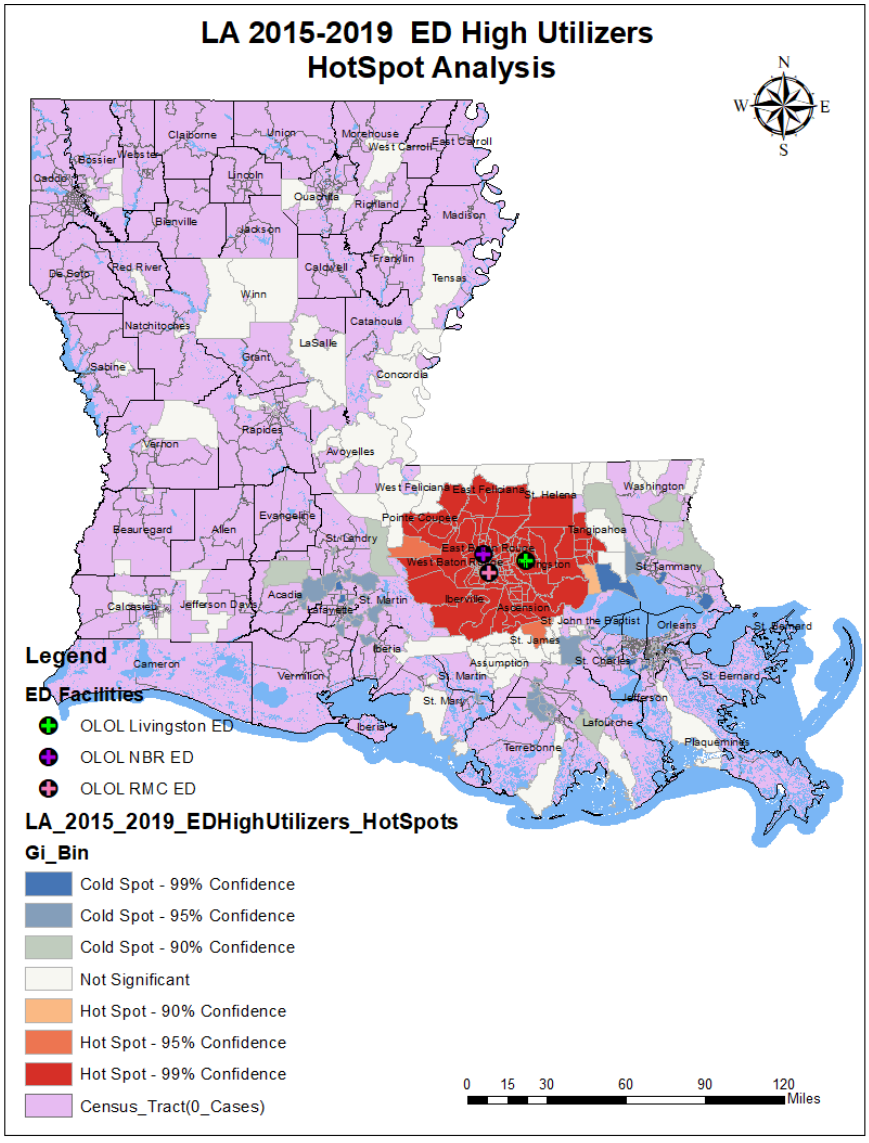


LA ED High Utilizers

Figure 4.1. Spatial Autocorrelation Report (Moran's I) for ED High Utilizers

4.4. Hot Spot Analysis

As discussed in Section 3.1.8, we used Spatial Statistics Tools->Mapping Clusters-> Hot Spots Analysis (Getis-Ord Gi). When the Hot Spots Analysis (Getis-Ord Gi) is run, the number of ED High Utilizers in a census tract (input feature) are analyzed within the context of neighboring input features. The clustering of high values (high number of ED High Utilizers) is considered as hot spot and the clustering of low values (low number of ED High Utilizers) is considered as cold spot. In this analysis we used unique ED High Utilizers (10,996) between 2015-2019 in LA.



LA ED High Utilizers

Figure 4.2. FMOLHS LA ED High Utilizers Hot Spot Analysis

To identify the Hot Spots of FMOLHS ED High Utilizers, the Hot Spot Analysis is run with the help of ArcGIS software on the count (number of ED High Utilizers) variable. We have considered the census tracts that have count value greater than 0 in the analysis. Also, the census tracts number,

parish and total ED High Utilizers count with the associated hot spot/ cold spot are shown in Appendix M. From Appendix M count values and Figure 4.2 we can observe that majority (more than 50%) of ED High Utilizer clusters of Hot Spots are in East Baton Rouge Parish. The other clusters of hot spots are in Iberville, East Feliciana, Livingston, Point Coupee, St. Helena, St. James, Tangipahoa, West Baton Rouge, West Feliciana Parishes. The Hot Spot map summary tables for ED High Utilizers are shown in Table 4.8. We also performed Cluster and Outlier analysis but we preferred Hot Spot analysis because as per the Cluster and Outlier analysis results shown in Appendix O, the number of ED High Utilizers under" Not Significant" category are 11% higher when compared to Hot Spot analysis results shown in Table 4.8. The maps generated as part of Cluster and Outlier analysis are shown in Appendix N, , average values of SVI and ACS Data in Appendix P and, the census tracts number, Parish and total ED High Utilizers count with the associated Cluster and Outlier's are shown in Appendix Q.

Table 4.8. Map Summary Table of FMOLHS LA ED High Utilizers

Results	Area (Kilometer Square)	# of Census Tracts	# of ED High Utilizers	Total Population	# ED High Utilizers/total pop x 1000
Hot Spot 99% Confidence	8557.3698	155	10648	878,553	12.12
Hot Spot 95% Confidence	451.8844	2	30	6,449	4.65
Hot Spot 90% Confidence	137.8266	1	1	5,342	0.19
Not Significant	19799.1608	74	193	375,097	0.51
Cold Spot 90% Confidence	2103.1884	9	13	42,480	0.31
Cold Spot 95% Confidence	1509.4957	33	41	200,727	0.20
Cold Spot 99% Confidence	592.7914	53	69	211,838	0.33

Table 4.9 shows the percentage averages of SVI and ACS data based on Hot spot Analysis results. In this analysis, we are not considering the "Not Significant "results because it is such a large area, and there is large amount of variability in the area (kilometer square) of Louisiana state. In comparison to the cold spots the hots spot have higher averages of % Age17, % Disability, % Single Parent, % Mobile Homes, % Crowded Housing, % Institutionalized Group Quarters, % Household Composition Rank, % Overall Rank SVI variable averages are higher in hot spot results. Additionally, the Per Capita Income average value is low in Hot Spots when compared to Cold Spots.

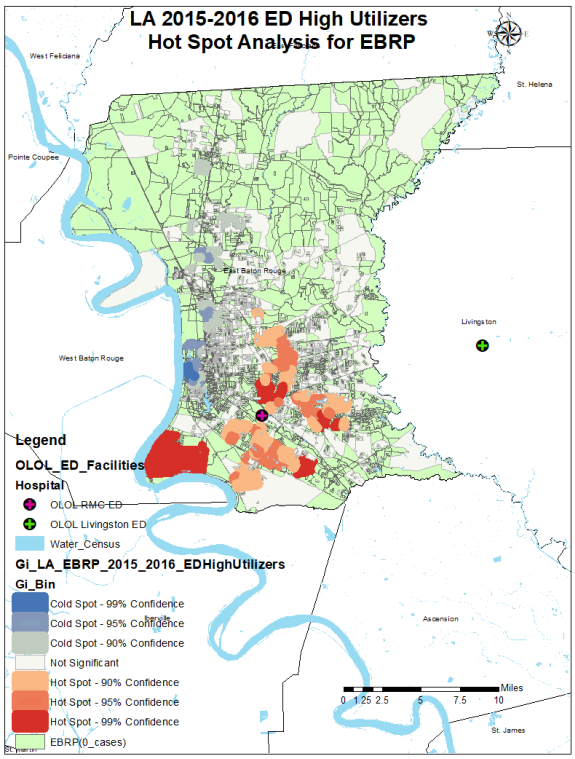
From Appendix M count values and Figure 4.2 we noticed that more than 50% of ED High Utilizer clusters of Hot Spots are in EBRP. So, we are performing Hot Spot Analysis with the help of ArcGIS software on the count (number of ED High Utilizers) variable for East Baton Rouge, Parish. From Figure 4.3 we can see the hot spots among the EBRP in the years 2015-2016, 2017, 2018 and, 2019.

In Figure 4.3 map (a) is shown for 2015-2016. After 2016 flood in LA many low-income people moved to student areas Burbank and Bluebonnet. Moreover, due to the Medicaid expansion in 2016 calendar year we have large number of ED high utilizers from low income neighborhoods when compared to other areas. From map (b) we can see that fair number of hot spots went away in 2017 due to the inauguration of the OLOL NBR ED. In the 2018 calendar year, there is 8.4% increase in the Medicaid enrollment for the LA population when compared to the Medicaid enrollment in 2017 [87] and from map (c) we can notice that there is an increase in the number of hot spots. In contrast, in 2019 there is 6.1% decrease in the Medicaid enrollment of LA population when compared to the 2018 calendar year [87]. Also, in 2019 the St. Elizabeth ED in Ascension parish was acquired by FMOLHS. Per map (d) notice that the hot spot confidence percentage is decreased when compared to 2018 because the census blocks have fewer number of ED High utilizers visiting the FMOLHS ED in 2019.

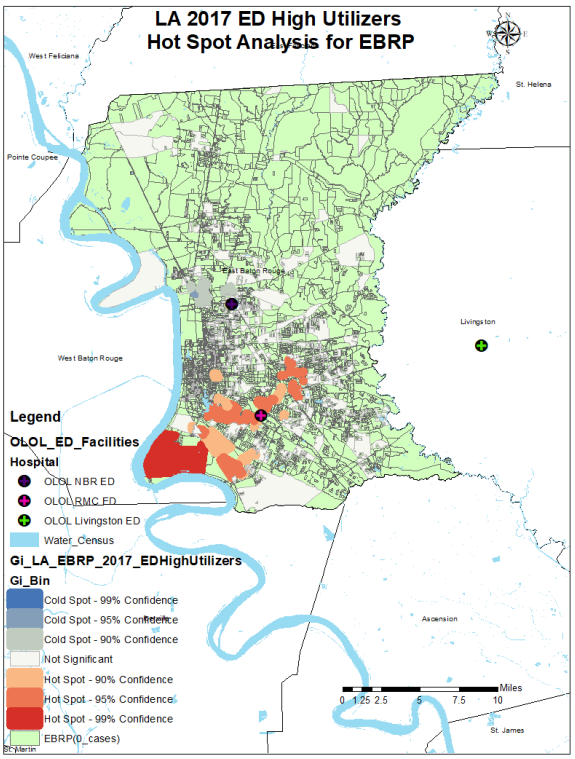
Table 4.9. Hot Spot Analysis Percentage Average Values of SVI and ACS Data Based on FMOLHS LA ED High Utilizers

LEGEND
High Average Values

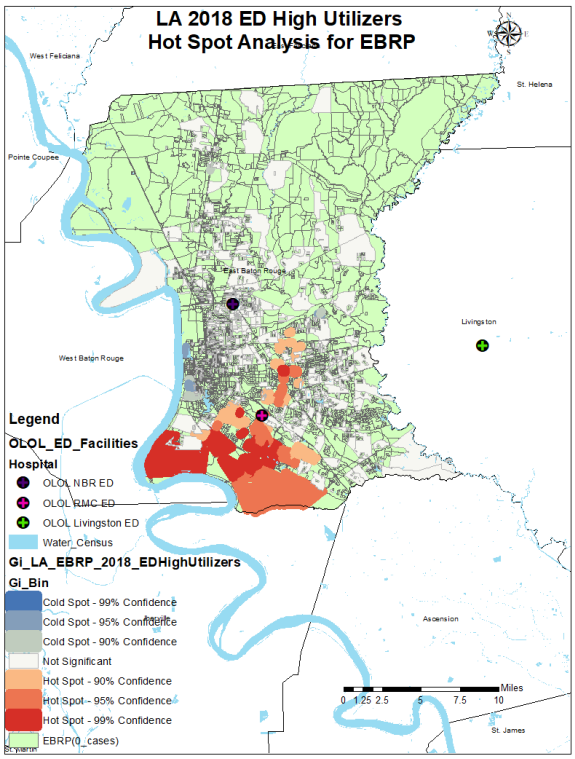
Results Confidence Spots/ SVI data % averages	Hot 99%	Hot 95%	Hot 90%	Cold 90%	Cold 95%	Cold 99%
LA Census Tract Data						
Poverty	19.35	20.95	19.30	18.73	20.35	22.85
Unemployment	7.41	5.60	7.10	6.48	6.71	8.81
Per Capita Income	28359.80	24826.00	25085.00	26501.56	26896.94	26415.79
No High School Dip	13.40	23.35	10.50	19.28	17.98	16.69
Age 65	13.67	15.10	14.00	16.89	13.25	14.44
Age 17	23.25	27.75	26.30	22.50	25.97	21.50
Disability	15.77	23.30	17.70	16.67	14.11	14.70
Single Parent	11.17	9.60	16.20	7.63	12.63	12.11
Minority	48.96	40.80	29.40	23.36	37.69	63.39
Limited English Proficiency	1.37	0.10	0.60	0.83	1.58	2.41
Housing 10 or More	10.73	0.00	0.50	2.36	5.63	12.22
Mobile Homes	11.00	24.30	8.50	19.41	15.80	2.58
Crowded Housing	2.81	3.55	2.90	1.92	2.77	1.82
No Vehicle	8.05	6.95	1.80	7.37	7.55	16.01
Institutionalized Group Quarters	2.96	1.85	0.10	0.97	1.28	2.55
Socioeconomic Rank	0.43	0.57	0.43	0.46	0.49	0.56
Household Composition Rank	0.47	0.79	0.85	0.49	0.51	0.45
Minority/Language Rank	0.51	0.35	0.47	0.34	0.54	0.69
Housing/Transportation Rank	0.52	0.51	0.40	0.43	0.60	0.40
Overall Rank	0.47	0.57	0.53	0.42	0.55	0.52
Uninsured	10.55	10.20	6.80	11.96	10.77	12.09



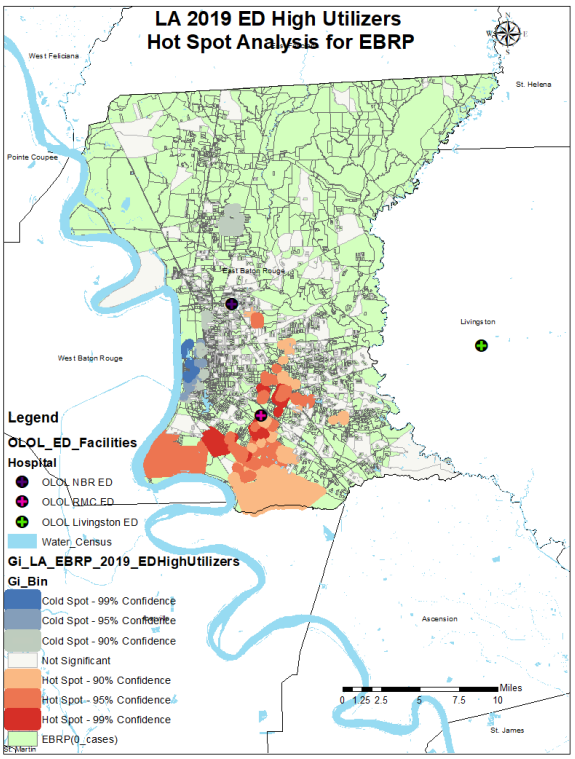
a) 2015-2016



b) 2017



c) 2018



d) 2019

Figure 4.3. FMOLHS EBRP ED High Utilizers Hot Spot Analysis in the year a)2015-2016, b)2017, c)2018, d)2019

4.5. Discussion

The primary outcomes of ED predictive classification model developments and spatial analysis are discussed in this section.

In the FMOLHS ED population visit information dataset, we had high number of Not-ED high utilizers and a smaller number of ED high utilizers. Because of the unbalanced data, the ED Predictive classification model development focused on identifying the ED high utilizers rather than Not-ED high utilizers. As we concentrated on the ED high utilizers(positives), the model development focused on the higher score of Accuracy, F1-Score, Recall, and Precision. The XGBoost classification model had the best Accuracy score of approximately 87.87% and is 0.42%, 1.05% higher when compared to Neural Network and Random Forest classification models respectively. Moreover, the evaluation metric scores for XGBoost model are: F1-Score is 87.36%, ROC_AUC is 95.11%, Recall, Precision, and Specificity are 0.87, 0.88, 0.88 respectively. The second-best model is Neural Network predictive classification model. This model evaluation metrics are Accuracy 87.45%, F1-Score is 87.34%, ROC_AUC is 95.01%, Recall, Precision, and Specificity are 0.87, 0.88, 0.88 respectively. These models will be useful for the FMOLHS to assess how changes in the ED patient visit data alter the FMOLHS ED high utilizers prediction based on previous patient data. The evaluation metrics on the validation showed that the on an average the developed predictive classification models had got only 12% of false positives and 13% of false negatives.

The ED predictive classification model has been validated on the unbalanced general FMOLHS ED visit data. The Neural Network classification model showed an Accuracy score of approximately 50.78% and is 0.53%,1.08% higher when compared to XGBoost and Random Forest classification models respectively. Moreover, the evaluation metrics of Neural Network model are: F1-Score is 2.79%, ROC_AUC is 73.89%, Recall, Precision, and Specificity are 0.86, 0.01, 0.50 respectively. Though the Neural network model had showed higher accuracy score, the evaluation metric Recall (0.87) has showed slightly higher scores in the XGBoost classification model. The evaluation metrics on the general population showed that the on an average the developed predictive classification models had identified large number (49%) of false positives and smaller number (14%) of false negatives.

Good results were achieved for the true positive rate and the model was able to classify 87 actual ED High utilizers out of 100 but overall the accuracy (50.25%) was poor due to many false positives, indicating additional work will be needed before the model is useful in practice.

To examine possible causes for the false positives, the “Admit” variable that differentiates the “ED Discharge and Admitted” patients was removed from the dataset and reran the model to identify if the classification model was relying too heavily on that one variable. Results from the classification model showed a 14% reduction in accuracy, true negative rate, and 3% reduction in true positive rate by removing the “Admit” variable. This implies that the SDOH and demographic factors account for only a small proportion of classification and “Admit” variable accounts for majority of the classification. With this change in the accuracy we noticed that the “Admit” variable should be included in the model and also that we are missing key information about the patients that would be needed to be able to accurately predict. The existing Epic readmission predictive model in OLOL has some overlap with the features used in our classification model but

then includes features concentrating on diagnosis and lab result codes. Moreover, the Epic model results are fairly close to the results we achieved in our study. In future research, by collecting more SDOH information at the patient level (rather than census level) and combining the diagnosis information, lab results, medications (used by the Epic model) to the ED visit patient data then the performance of the model likely can be significantly improved.

The hot spot analysis results will be useful for the healthcare organization/ healthcare advocates to guide the people to get connected with right resources in order to improve their health based on the features associated with the census blocks that have high number of ED high utilizers. For example, the ED high utilizers hot spots showed limited English proficiency and the patient advocates can use this result to guide the ED high utilizers to various health promotion programs across EBRP that can benefit the individual to improve their health and at the same time educating the people to visit other alternatives such as primary care, urgent care in order to reduce the ED visits for non-emergency conditions.

The hotspot analysis can be improved by including more patients and more hospitals. For example, not just the people going to OLOL facilities but including all the people who visit regional hospitals not affiliated to FMOLHS. In future research, we can also perform the hotspot analysis based on SDOH factors such as per capita rate or population density by breaking down to the smaller geographical areas. This can help us to know if the less densely populated areas are having larger number of ED high utilizers due to the impact of SDOH factors.

The key points of hot spot analysis are:

- The census tracts that have higher percentages of SVI variables such as poverty, unemployment, no high school diploma, crowded housing, single parent, elderly population, institutionalized group quarters, disability, minority, multiunit housing are highly correlated for the increase in ED High Utilizers among the FMOLHS patient population.
- East Baton Rouge parish is considered as socially vulnerable area and people living in these areas utilize ED more for non-emergency reasons.
    - The Medicaid expansion in 2016 and inauguration of OLOL NBR ED in 2017 led to the FMOLHS ED High Utilizers in EBRP between 2015-2017.
    - In 2018, for the first time we noticed the hot spots in the West Baton Rouge parish due to increase in Medicaid enrollment across LA state.
    - In the 2019 calendar year there is a decrease in the Medicaid enrollment in the LA population and parallelly St. Elizabeth ED in Ascension parish was acquired by FMOLHS and went live in EPIC. So, the majority of hot spots are from Ascension, Livingston parishes but, the EBRP showed cold spots.
    - In comparison to the cold spots, the hot spots have higher averages of single parent population; limited English proficiency; people living in multiunit houses; minority status population; people living in crowded housing; are higher in hot spot results. Additionally, the Per Capita Income average value is low in hot spots when compared to cold spots. These factors in East Baton Rouge parish might drive the people to use the ED for non-emergency reasons.

- The FMOLHS ED management might work on developing a screening tool that will improve the existing process of integrating the SODH information to the patient information during/after their visit.
- The FMOLHS management have patients coming from census tracts that have higher percentages of poverty rates, unemployment rates, single parent, disability, minority. An interactive web application can be developed to provide necessary social prescriptions such as healthy nutrient food outlets, access to exercise, and guidance on financial situation. This might help to reduce the financial burden caused by ED over-utilization at FMOLHS.

Chapter 5. COVID Retrospective Study Results and Analysis

5.1. Retrospective Study of FMOLHS COVID Tested Patients

The purpose of this study is to identify the characteristics of COVID Positive tested patients and identify factors that may have contributed to the outbreak of the diseases across LA based on FMOLHS COVID patient data. In this study, COVID tested patient's information (18 and above age) between March 13th 2020 to July 15th 2020 are used for the analysis. To identify the patient demographics, clinical characteristics, SVI data and ACS occupational census data that are highly responsible for the COVID-19 disease the following retrospective studies have been conducted.

5.1.1. Correlation Analysis on COVID Tested Patients

As discussed in Section 3.1.7.2, the institutionalized patient population are treated as a separate population because they do not reflect the social and demographic characteristics of the neighborhoods their institution is located in. To understand the influence of SVI data, patient data discussed in Table 3.8 with respect to the COVID result ('Positive/Negative') we are considering the 2204 records of non-institutionalized positive patients. These records are added to the randomly selected 2204 negative non-institutionalized patients using SPSS Data->Select Cases->Random sample of cases. This dataset has satisfied the Spearman correlation test assumptions. The Spearman correlation is done between the *numeric(continuous/binary)* dependent and independent variables. The H_0 (null hypothesis) is defined as "There is no[monotonic] association between the variables"
The following SPSS code is used to run the Spearman correlation analysis.

```
NONPAR CORR
  /VARIABLES=BMI Age LOS Living_Status Sex COVID_results
Charlson_predictive_mortality E_TOTPOP
    EP_POV EP_UNEMP EP_PCI EP_NOHSDP EP_AGE65 EP_AGE17
EP_DISABL EP_SNGPNT EP_MINRTY EP_LIMENG EP_MUNIT
    EP_MOBILE EP_CROWD EP_NOVEH EP_GROUPQ RPL_THEME1
RPL_THEME2 RPL_THEME3 RPL_THEME4 RPL_THEMES
    EP_UNINSUR EP_OCC_MBSA EP_OCC_SER EP_OCC_SAL_OFF
EP_OCC_NRCM EP_OCC_PTMM Auto_immun CEVD CKD CLD
    CVD Diabetes ESRD Hep_B Hep_C HIV HTN Immunocompromised
Obesity overweight PVD Respiratory
  /PRINT=SPEARMAN TWOTAIL NOSIG
  /MISSING=PAIRWISE.
```

The results of the Spearman Correlation analysis are shown in Table 5.1. The correlation scores among other variables are shown in Appendix R.

From Table 5.1 we can see that BMI, Age, LOS, Living Status, EP_UNEMP, EP_SNGPNT, EP_MINRTY, RPL_THEME3, RPL_THEMES, EP_OCC_SER, Obesity have weak correlation scores and the rest of the factors have correlation scores less than 0.1(irrespective of + or -) with respect to COVID_Results. Additionally, Age, LOS, COVID_results, CPMI, CEVD, CKD, CVD,

Diabetes, ESRD, HTN have weak correlation scores and the rest of the factors have correlation scores less than 0.1 with respect to Living Status.

Table 5.1. Spearman Correlation Scores Between Numeric Variables

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	Living Status		Covid Result	
	(r_s)	p-value	(r_s)	p-value
BMI	-0.01	0.711	0.14	0.000
Age	0.19	0.000	-0.10	0.000
LOS	0.18	0.000	0.19	0.000
Living_Status	***		0.12	0.000
Sex	0.07	0.000	-0.02	0.200
COVID_results	0.12	0.000	***	
CPMI	-0.24	0.000	0.08	0.000
E_TOTPOP	-0.04	0.024	-0.01	0.392
EP_POV	0.04	0.004	0.08	0.000
EP_UNEMP	0.03	0.068	0.10	0.000
EP_PCI	-0.05	0.002	-0.08	0.000
EP_NOHSDP	0.04	0.008	0.06	0.000
EP_AGE65	0.01	0.714	-0.05	0.002
EP_AGE17	0.00	0.885	0.05	0.001
EP_DISABL	0.03	0.102	0.03	0.037
EP_SNGPNT	0.03	0.061	0.10	0.000
EP_MINRTY	0.06	0.000	0.19	0.000
EP_LIMENG	0.02	0.123	0.05	0.001
EP_MUNIT	-0.01	0.545	0.03	0.098
EP_MOBILE	-0.02	0.268	-0.08	0.000
EP_CROWD	0.03	0.056	0.03	0.033
EP_NOVEH	0.04	0.014	0.09	0.000
EP_GROUPQ	-0.01	0.609	0.00	0.796
RPL_THEME1	0.05	0.003	0.09	0.000
RPL_THEME2	0.03	0.061	0.07	0.000
RPL_THEME3	0.06	0.000	0.17	0.000
RPL_THEME4	0.01	0.459	0.03	0.042
RPL_THEMES	0.05	0.003	0.11	0.000
EP_UNINSUR	0.03	0.093	0.07	0.000
EP_OCC_MBSA	-0.06	0.000	-0.08	0.000
EP_OCC_SER	0.05	0.003	0.10	0.000
EP_OCC_SAL_OFF	-0.02	0.159	-0.02	0.290
EP_OCC_NRCM	0.02	0.266	-0.03	0.033
EP_OCC_PTMM	0.03	0.030	0.06	0.000

(Table Cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	Living Status		Covid Result	
	r_s	p-value	r_s	p-value
Auto_immun	0.01	0.589	-0.05	0.001
CEVD	0.11	0.000	0.02	0.296
CKD	0.16	0.000	0.01	0.658
CLD	0.02	0.240	-0.01	0.493
CVD	0.13	0.000	-0.01	0.754
Diabetes	0.14	0.000	0.07	0.000
ESRD	0.11	0.000	0.04	0.020
Hep_B	0.02	0.280	0.00	1.000
Hep_C	0.02	0.125	-0.04	0.007
HIV	0.03	0.024	0.00	0.780
HTN	0.12	0.000	-0.02	0.140
Immunocompromised	0.01	0.690	-0.01	0.592
Obesity	0.09	0.000	0.19	0.000
overweight	0.01	0.713	0.03	0.033
PVD	0.05	0.002	0.00	0.921
Respiratory	0.06	0.000	-0.07	0.000

To find the correlation scores between *numeric* (continuous/binary) dependent and *categorical* independent variables an Eta Coefficient test was run. In this scenario the *null hypothesis* is defined as H_0 : There will not be any association between two variables. The results are shown in Table 5.2.

The following SPSS Code is used to find the association between the Living Status, COVID_results and other factors using Eta Coefficient test.

```
CROSSTABS
  /TABLES=Type Smoking_status Race_Ethnicity Prim_RUCA_ BY
Living_Status COVID_results
  /FORMAT=AVALUE TABLES
  /STATISTICS=ETA
  /CELLS=COUNT
  /COUNT ROUND CELL
```

Table 5.2. Results of the Eta Coefficient Test

Eta ($\dot{\eta}$)- Nominal by Interval Measure		
Independent/Dependent	**COVID_Results**	**Living_Status**
Type	0.138	0.299
Smoking Status	0.220	0.093
Race_Ethnicity	0.296	0.060
Prim_RUCA	0.119	0.028

Table 5.2 shows:

- There is a weak association between Type and COVID_Results; the Eta coefficient $\dot{\eta}$ = 0.138 and $\dot{\eta}^2$ = 0.019 implies that we can reject null hypothesis and there is an association between Type and COVID_Results; the Type variable contributes to 1.9% of variance in COVID_Results
- There is a weak association between Smoking Status and COVID_Results; the Eta coefficient $\dot{\eta}$ = 0.220 and $\dot{\eta}^2$ = 0.048 implies that we can reject null hypothesis and there is an association between Smoking Status and COVID_Results; the Smoking Status variable contributes to 4.8% of variance in COVID_Results
- There is a weak association between Race_Ethnicity and COVID_Results; the Eta coefficient $\dot{\eta}$ = 0.296 and $\dot{\eta}^2$ = 0.087 implies that we can reject null hypothesis and there is an association between Race_Ethnicity and COVID_Results; the Race_Ethnicity variable contributes to 8.8% of variance in COVID_Results
- There is a weak association between Prim_RUCA and COVID_Results; the Eta coefficient $\dot{\eta}$ = 0.119 and $\dot{\eta}^2$ = 0.014 implies that we can reject null hypothesis and there is an association between Prim_RUCA and COVID_Results; the Prim_RUCA variable contributes to 1.4% of variance in COVID_Results

Likewise, there is a weak association between Living Status and Type, and negligible association between Living Status and Smoking Status, Living Status and Race_Ethnicity, Living Status and Prim_RUCA. The variables Type, Smoking Status, Race_Ethnicity, and Prim_RUCA contribute 8.9%, 0.9%, 0.4%, and 0.1% of variance in Living Status respectively.

From Table 5.1 and Table 5.2 can observe that the COVID test result ('Positive/Negative') for the general patients (non-institutionalized FMOLHS patients) are correlated with patient demographics: Age, BMI, Race_Ethnicity, Smoking Status, clinical data such as LOS, Type, underlying conditions: Obesity, SVI data: EP_UMEMP, EP_SNGPNT, EP_MINRTY, RPL_THEME3, RPL_THEMES and, ACS data: EP_OCC_SER. This implies that FMOLHS patients who belongs to minority race/ethnicity, single parent, having occupations related to service such as plumber services, restaurants, computer services, high socioeconomic rank (RPL_THEME1) and overall rank (if RPL_THEME is high then that area have high social vulnerability) might have higher chance of acquiring COVID-19 disease.

By considering Living Status ('Alive/Deceased') for the general patients (non-institutionalized FMOLHS patients), it is influenced by the by the patient demographics: Age, clinical data: LOS, CPMI, underlying conditions: CEVD, CKD, CVD, Diabetes, ESRD, HTN.

5.1.2. Comparison of FMOLHS COVID Positive and Negative Tested Patients

Based on the data in the tables shown in Appendix S, Appendix T, Appendix U, Appendix V, Appendix W, and Appendix X we can interpret the differences and similarities between positive and negative tested COVID FMOLHS patients.

- Out of 16472 tested patients, 13714 (83.26%) are negative and 2758 (16.74%) are positive.

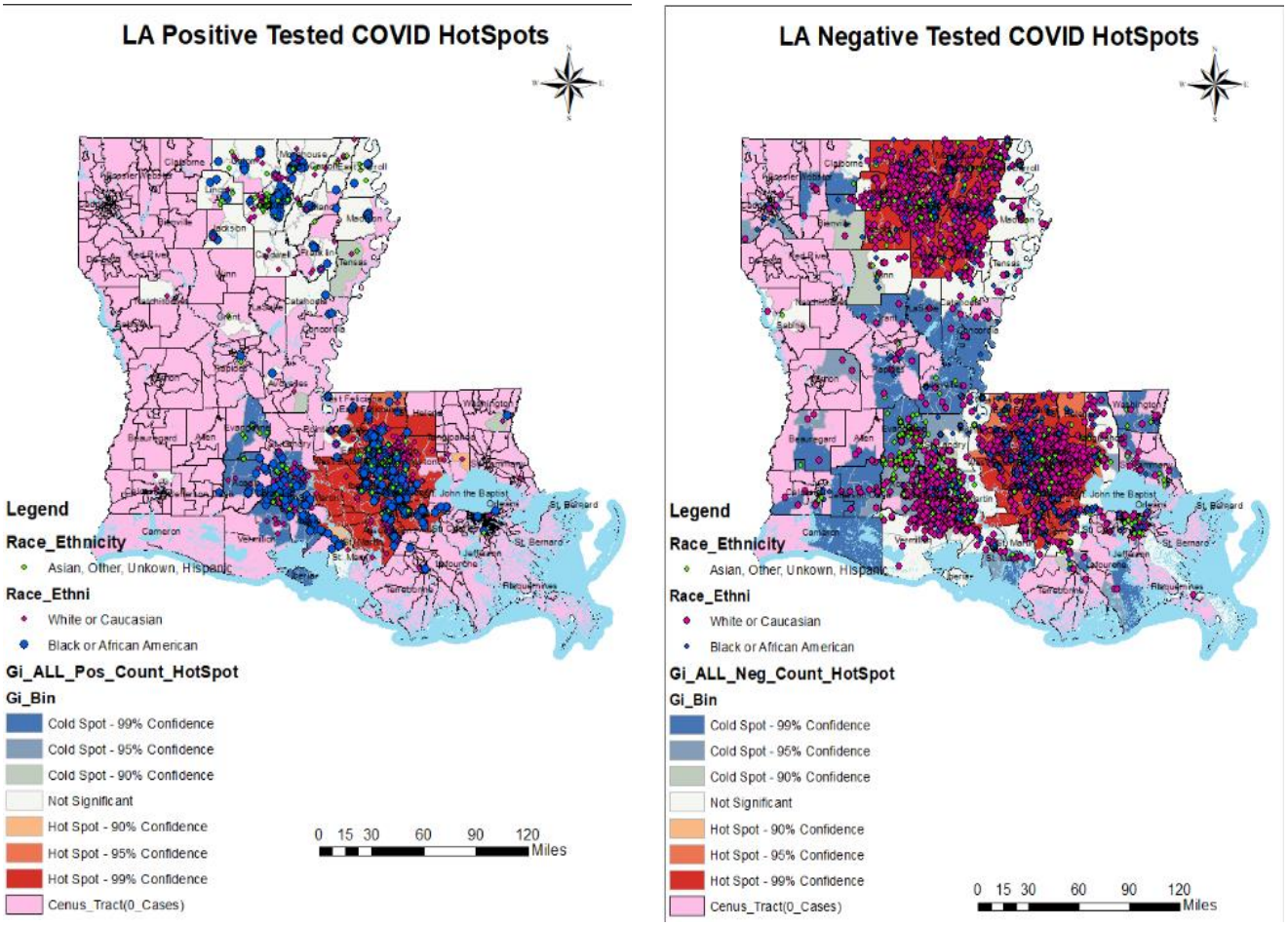
- Out of the total COVID tested patients, 3.65% died. Of this,1.98% died due to underlying health conditions: HTN, Diabetes, CVD, CKD, CEVD, Obesity, Respiratory, other problems and 1.67% died due to COVID-19 disease.
- The mortality rate is higher in males when compared to females in both COVID positive and negative tested patients (Appendix S, Appendix V).
- The average age and CPMI of FMOLHS patients who tested positive for COVID-19 and died is approximately 72 and 36.26% respectively, and for positive patients who are alive is approximately 54 and 57.94% (Appendix S). However, for the patients who tested negative for COVID-19 and died is approximately around 70, 32.17% and negative patients who are alive is around 56, 50.47% (Appendix V).
- We could notice higher percentage (47.65%) of negative tested FMOLHS patients with Smoking Status as "Current" and "Former" when compared to positive tested patients (28.21%) (Appendix S, Appendix V).
- There is higher rate (8.23%) of death rates among FMOLHS COVID positive patients admitted in the hospital when compared to the negative tested patients (1.95%) admitted in the hospital (Appendix S, Appendix V).
- 20.09% of all positive tested patients are institutionalized patients with average age around 74 and 4.06% of them are deceased. 8.42% of all negative tested patients are institutionalized patients with average age around 73 and 0.36% of them are deceased. Also, most of the institutionalized patients are coming from Nursing Homes (Appendix S, Appendix V).
- In both positive and negative cases, the underlying health conditions are similar and they have high percentages of HTN, Diabetes, CVD, Respiratory, Obesity, CKD, CEVD and, the rest of the problem categories such as PVD, Overweight, Immunocomp, ESRD, Hep B, Hep C, HIV, CLD, Auto immune are having negligible percentages among COVID tested patients. Moreover, HTN, Diabetes and CVD are the top 3 underlying conditions for both positive and negative tested patients (Appendix T, Appendix W).
- The Average_BMI is higher in positive patients when compared to negative tested patients (Appendix S, Appendix V).
- The negative tested patients have lower average percentage values of SVI data when compared to positive tested patients. For example, EP_PCI is low for positive patient, EP_SNGPNT (11.53%), EP_MINRTY (46.54%), EP_LIMENG (1.25%), EP_MUNIT (7.87%), EP_GROUPQ (3.46%), RPLTHEME (0.46) (Overall ranking based on SVI data) are high in positive patients. The overall ranking indicates that the population who have high social vulnerability are at high risk of getting COVID-19 when compared to the counterparts (Appendix U, Appendix X).
- FMOLHS patients who have occupations related to service (18.99%) are at higher risk of acquiring COVID-19 disease when compared to counterparts. For example, jobs such as plumber services, restaurants, computer services etc... (Appendix U, Appendix X).
- In total 16472 tested patients, 6141 (37.28%) are black, 8849 (53.72%) are white but, 1553 (9.43%) of all tested are positive and black. This implies that approximately 77.71%

of positive tested (2758) people are blacks. Moreover, 7632 (46.33%) of all tested patients are white and negative (Appendix S, Appendix V).

Based on the above discussion, we noticed that the positive tested population groups are different from negative tested population. This is because the positive tested group has elderly population staying in socially vulnerable areas with underlying health conditions and employed at service occupations (e.g., grocery stores, restaurants etc.)

5.1.2.1. Hot Spot Analysis

As discussed in Section 3.1.8, we used Spatial Statistics Tools->Mapping Clusters-> Hot Spots Analysis (Getis-Ord Gi) in ArcGIS. When the Hot Spots Analysis (Getis-Ord Gi) is run, the number of COVID positive cases in a census tract (input feature) are analyzed within the context of neighboring input features. The clustering of high values (high number of COVID Positive patients or high number of COVID Negative patients) is considered as hot spot and the clustering of low values (small number of COVID Positive patients or small number of COVID Negative patients) is considered as cold spot.



a) Positive b) Negative

Figure 5.1. FMOLHS COVID a) Positive and b) Negative Tested Patients Hot Spot Analysis

To identify the Hot Spots of FMOLHS COVID Positive and Negative tested patients the Hot Spot Analysis is run with the help of ArcGIS software on the count greater than 0 (number of positive/number of negative patients) variable. From Figure 5.1 map (a) we can observe that census tracts with less/ no COVID positive patients shown in map has more hot spots/ clusters of negative tested patients in map (b). Figure 5.1 shows that the black ethnicity patients are more in Positive tested patients. In contrary, the White ethnicity patients are more in Negative tested patients.

5.1.3. Comparison of FMOLHS COVID Positive Institutionalized and Non-Institutionalized Patients

Based on the tables shown in Appendix S, Appendix Y, Appendix Z, Appendix AA, Appendix AB, and Appendix AC we can identify the differences and similarities between positive institutionalized and non-institutionalized COVID FMOLHS patients.

- Out of 2758 positive tested patients, 2204 (79.91%) are non-institutionalized and 554 (20.09%) are institutionalized (Appendix S).
- Out of total COVID positive tested patients, 9.97% are dead. In this 5.91% are non-institutionalized and 4.06% are institutionalized (Appendix S).
- Though the females are tested positive at higher rate than males among institutionalized and non-institutionalized patients, the mortality rate is high in males (Appendix Y, Appendix AA).
- The average age, CPMI of FMOLHS positive institutionalized patients who die due to COVID-19 is approximately around 75, 25.32 and positive institutionalized patients who are alive is around 73, 31.58. However, for positive non-institutionalized patients who die due to COVID-19 is approximately around 69, 43.78 and positive non-institutionalized patients who are alive is around 50, 63.65 (Appendix Y, Appendix AA).
- We could notice higher percentage (44.40%) of institutionalized FMOLHS patients with Smoking Status as “Current” and “Former” when compared to non-institutionalized patients (24.13%) (Appendix Y, Appendix AA).
- There is higher rate (14.80%) of death rates among institutionalized patients admitted in the hospital when compared to the non-institutionalized patients (6.40%) admitted in the hospital (Appendix Y, Appendix AA).
- In both institutionalized and non-institutionalized positive tested cases, the underlying problem are similar and they have high percentages of HTN, Diabetes, CVD, Respiratory, Obesity, CKD, CEVD and, the rest of the problem categories such as PVD, Overweight, Immunocomp, ESRD, Hep B, Hep C, HIV, CLD, Auto immune are having negligible percentages among COVID positive tested patients (Appendix Z, Appendix AB).
- HTN, Diabetes, CVD are the top 3 underlying health conditions for institutionalized patients (Appendix Z).
- HTN, Diabetes, Obesity are the top 3 underlying conditions for non-institutionalized patients (Appendix AB).

- The minority ethnicity population showed higher percentage of positive cases among non-institutionalized patients when compared to institutionalized patients (Appendix Y, Appendix AA).
- There were no deaths among Asian and Hispanic FMOLHS non-institutionalized patients (Appendix AA).
- The Average age is high in positive institutionalized patients when compared to positive non-institutionalized patients. Moreover, majority of the positive tested institutionalized people are of age above 60 (Appendix Y, Appendix AA).

Based on the above discussion, we noticed that the institutionalized positive patient population groups are different from non-institutionalized positive patient population. This is because the institutionalized positive patient group has elderly population with many underlying health conditions. Moreover, these group of people have experienced higher death rates when compared to non-institutionalized positive patient population.

5.1.3.1. Spatial Autocorrelation

As discussed in Section 3.1.8, to understand the patterns of institutionalized and non-institutionalized positive patients we used Spatial Statistics Tools->Analyzing Patterns-> Spatial Autocorrelation (Moran's I). To run this analysis, we have included the census tracts that have positive patients count greater than 0.

- We choose EUCLIDEAN distance method as the study area is small (Example: Census Tracts) and it calculates the distance between two points connected by a straight-line.
- We choose conceptualization of spatial relationships as INVERSE DISTANCE because the nearby features have a larger influence on the computations of a target feature than features that are far away.

The reports that are generated are shown in the Figure 5.2, the values are shown in Table 5.3

Table 5.3. Global Moran's I Results

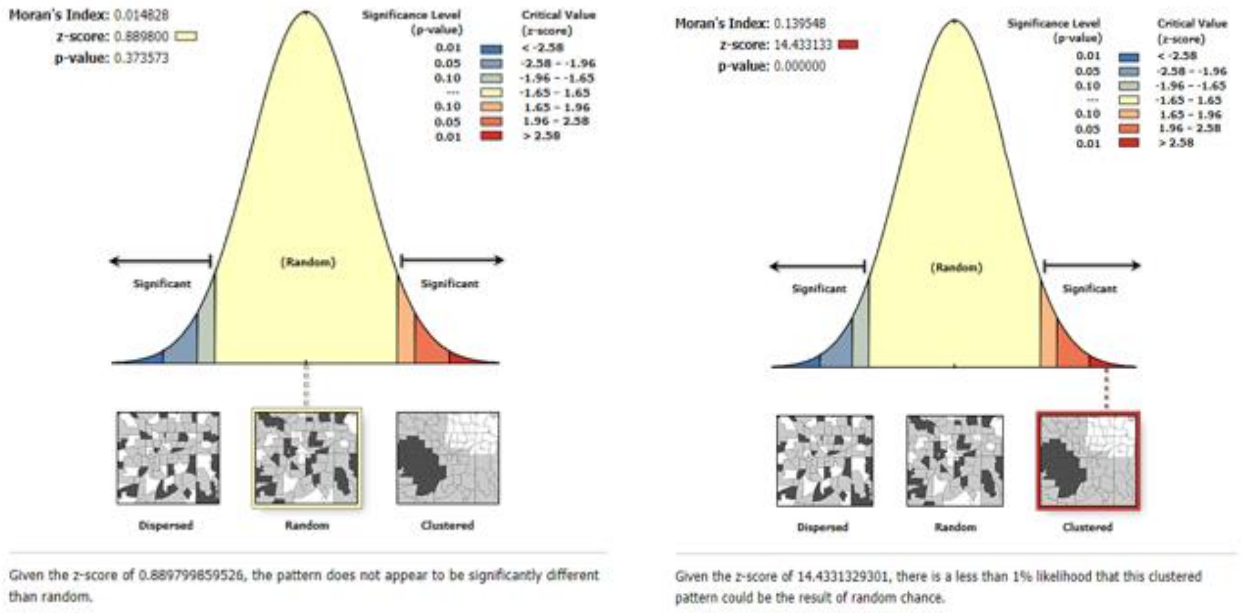
Global Moran's I Summary	Institutionalized	Non-Institutionalized
Moran's Index:	0.014828	0.139548
Expected Index:	-0.014286	-0.002755
Variance:	0.001071	0.000097
Z-score:	0.889800	14.433133
p-value	0.373573	0.000000

From Table 5.3 and Figure 5.2 we can see that

- The Variance is 0.001071 for institutionalized and 0.000097 for non-institutionalized positive patients
- P-value is a probability, it is used to check whether the observed spatial pattern is the result of random process and based on that we decided whether to reject or not to reject the null hypothesis. The null hypothesis in these scenarios is defined as "the values associated with features are randomly distributed".
    - Here P-value is 0.373573 (>0.05) and not statistically significant for institutionalized positive patients. So, null hypothesis cannot be rejected.
    - Here P-value is near 0 (<0.01) and statistically significant for non-institutionalized positive patients. So, null hypothesis can be rejected.

- Moran's Index varies between −1 and +1. A value near +1 indicates that the census tracts in the study are area clustered; and a value near −1 indicates that the census tracts in the study area are dispersed. If a Moran's I is close to 0, it indicates a random pattern or absence of spatial autocorrelation.
    - **Non-Institutionalized patients**: The Z-score for Moran's I is 14.433133 and p-value is 0.00, these values indicate that the detected spatial autocorrelation (positive non-institutionalized patients) was statistically significant. In this scenario, Moran's I is 0.139548, it indicates that the positive non-institutionalized patient census tracts in the study area are clustered.
    - **Institutionalized patients:** Z-score for Moran's I is 0.889800 and p-value is 0.373573, these values indicate that the detected spatial autocorrelation (positive institutionalized patients) was not statistically significant. In this analysis, the Moran's I's is 0.014828 and close to zero. This indicates the absence of spatial autocorrelation

As discussed in earlier Section 3.1.7.2, the institutionalized patients should be treated as separate population and it is seen in the spatial autocorrelation analysis that the FMOLHS institutionalized positive patients are not clustered but randomly distributed. So, the institutionalized population do not reflect the social and demographic characteristics of the neighborhoods their institution is located in.



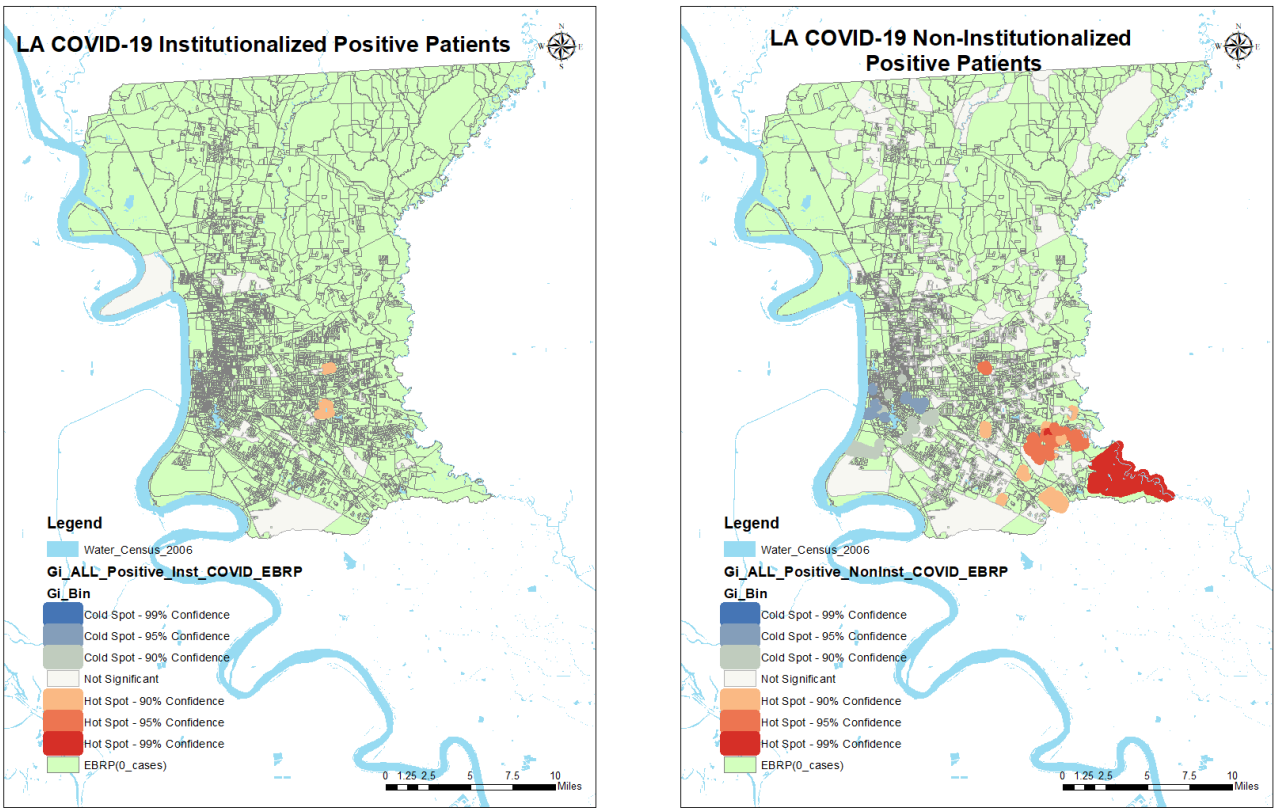
a) Institutionalized Patients b) Non-Institutionalized

Figure 5.2 Spatial Autocorrelation Report (Moran's I) for a) Institutionalized Patients b) Non-Institutionalized Patients

5.1.3.2 Hot Spot Analysis

As discussed in Section 3.1.8, we used Spatial Statistics Tools->Mapping Clusters-> Hot Spots Analysis (Getis-Ord Gi). When the Hot Spots Analysis (Getis-Ord Gi) is run for EBRP, the number

of COVID positive cases in a census tract (input feature) are analyzed within the context of neighboring input features. The clustering of high values (high number of COVID Positive institutionalized patients or high number of COVID Positive non-institutionalized patients) is considered as hot spot and the clustering of low values (small number of COVID Positive institutionalized patients or small number of COVID Positive non-institutionalized patients) is considered as cold spot.



a) Institutionalized b) Non-Institutionalized

Figure 5.3. FMOLHS EBRP COVID a) Institutionalized and b) Non-Institutionalized Positive Tested Patients Hot Spot Analysis

To identify the Hot Spots of FMOLHS COVID Positive institutionalized and non-institutionalized patients, the Hot Spot Analysis is run with the help of ArcGIS software on the count (number of positive institutionalized/number of negative institutionalized) variable. We have considered the census tracts that have count value greater than 0 in the analysis. From Figure 5.3 map a we can observe that there are only few census tracts with COVID positive institutionalized patients identified as hot spots. In contrary, we have more hot spots/ clusters of non-institutionalized positive tested patients in map b. By looking and comparing both the maps we can notice that most of the cold spots and hot spots in the non-institutionalized patients map are in the census tract that does not have the institutionalized patients hot or cold spots. This indicates that institutionalized population are different from the non-institutionalized (general) patients of FMOLHS.

The Hot Spot map summary tables for Institutionalized and Non-Institutionalized positive patients are shown in Table 5.4 and Table 5.5. In the Table 5.5 and Figure 5.3 map b we could see

clusters of non-institutionalized COVID-19 positive cases in Northwest and South-Central Louisiana, and compromised of 196 census tracts that are identified as Hot Spots. Also, the census tracts number, parish and total positive patients count with the associated hot spot/ cold spot are shown in Appendix AD and Appendix AE. We also performed Cluster and Outlier analysis but we preferred Hot Spot analysis because the Cluster and Outlier analysis has considered most of the positive patients as" Not Significant" when compared to Hot Spot analysis. The maps generated as part of Cluster and Outlier analysis are shown in Appendix AF, map summary tables in Appendix AG , average values of SVI and ACS Data in Appendix AH and, the census tracts number, Parish and total positive patients count with the associated Cluster and Outlier's are shown in Appendix AI and Appendix AJ.

Table 5.6 shows the percentage averages of SVI and ACS data based on Hot spot Analysis results. In this analysis, we are not considering the "Not Significant "results because it is such a large area, and there is large amount of variability in the area (kilometer square) of Louisiana state. In Table 5.6 we can see that the % Housing with 10 or More Units, % Mobile Homes, % of Management, Business, Science & Art Occupations averages are high in hot spot results when compared to cold spots.

Table 5.4. Map Summary Table of Institutionalized Positive Patients

Results	Area (Kilometer Square)	# of Census Tracts	# of COVID-19 Positive	Total Population	# COVID-19/total pop x 1000
Hot Spot 99% Confidence	40.77	1	15	6,742	2.22
Hot Spot 95% Confidence	1053.24	23	277	141,894	1.95
Hot Spot 90% Confidence	185.44	7	100	53,793	1.86
Not Significant	4035.73	40	162	171,823	0.94

Table 5.5. Map Summary Table of Non-Institutionalized Positive Patients

Results	Area (Kilometer Square)	# of Census Tracts	# of COVID-19 Positive	Total Population	# COVID-19/total pop x 1000
Hot Spot 99% Confidence	9841.38	161	1308	884,022	1.48
Hot Spot 95% Confidence	5277.40	19	69	73,665	0.94
Hot Spot 90% Confidence	4294.21	16	89	80,401	1.11
Not Significant	9080.78	65	432	244,422	1.77
Cold Spot 90% Confidence	2547.83	4	5	16,054	0.31
Cold Spot 95% Confidence	310.94	9	10	39,216	0.25
Cold Spot 99% Confidence	6193.16	90	290	486,767	0.60

Table 5.6. Hot Spot Analysis Percentage Average Values of SVI and ACS Data

LEGEND

High Average Values

Results Confidence Spots/ SVI data % averages	Hot 99%	Hot 95%	Hot 90%	Cold 90%	Cold 95%	Cold 99%
Poverty	18.68	26.48	21.31	27.23	24.71	22.17
Unemployment	7.20	6.58	6.18	6.88	9.44	7.34
Per Capita Income	28287	21214	22739	18058	22183	24328
No High School Dip	13.87	19.54	17.51	21.75	17.03	19.46
Age 65	14.00	16.47	16.76	17.85	13.53	13.70
Age 17	23.29	23.37	23.48	21.95	26.86	25.12
Disability	15.54	16.34	14.97	19.27	13.83	15.25
Single Parent	11.47	10.71	8.97	9.00	15.40	12.03
Minority	49.62	44.17	31.63	26.27	62.53	39.83
Limited English Proficiency	1.36	0.87	0.47	0.63	3.36	1.64
Housing 10 or More	9.78	2.29	0.68	0.80	4.73	5.93
Mobile Homes	11.64	22.33	28.24	24.25	6.67	16.03
Crowded Housing	2.82	1.92	2.45	2.95	3.08	3.05
No Vehicle	7.95	11.34	6.91	8.98	13.33	9.63
Institutionalized Group Quarters	2.43	2.95	4.97	6.77	0.18	2.12
Socioeconomic Rank	0.43	0.60	0.52	0.67	0.57	0.55
Household Composition Rank	0.48	0.57	0.45	0.54	0.64	0.53
Minority/Language Rank	0.51	0.46	0.29	0.31	0.61	0.53
Housing/Transportation Rank	0.51	0.60	0.58	0.69	0.39	0.61
Overall Rank	0.47	0.59	0.47	0.61	0.56	0.58
Uninsured	10.34	12.63	13.16	15.23	14.01	12.02
Management, Business, Science & Art Occupations	32.81	26.06	28.62	22.28	27.76	28.75
Service Occupations	19.60	21.28	19.61	19.55	27.74	20.57
Sales & Office Occupations	22.12	20.32	22.66	18.00	19.96	22.69
Natural resources, construction, and maintenance occupations	10.82	15.11	13.75	25.50	10.22	13.10
Production, transportation, and material moving occupations	14.61	17.33	15.38	14.68	14.41	14.87

5.2. Discussion

In this section, the primary outcomes of the analysis conducted on the FMOLHS COVID-19 tested patients are discussed.

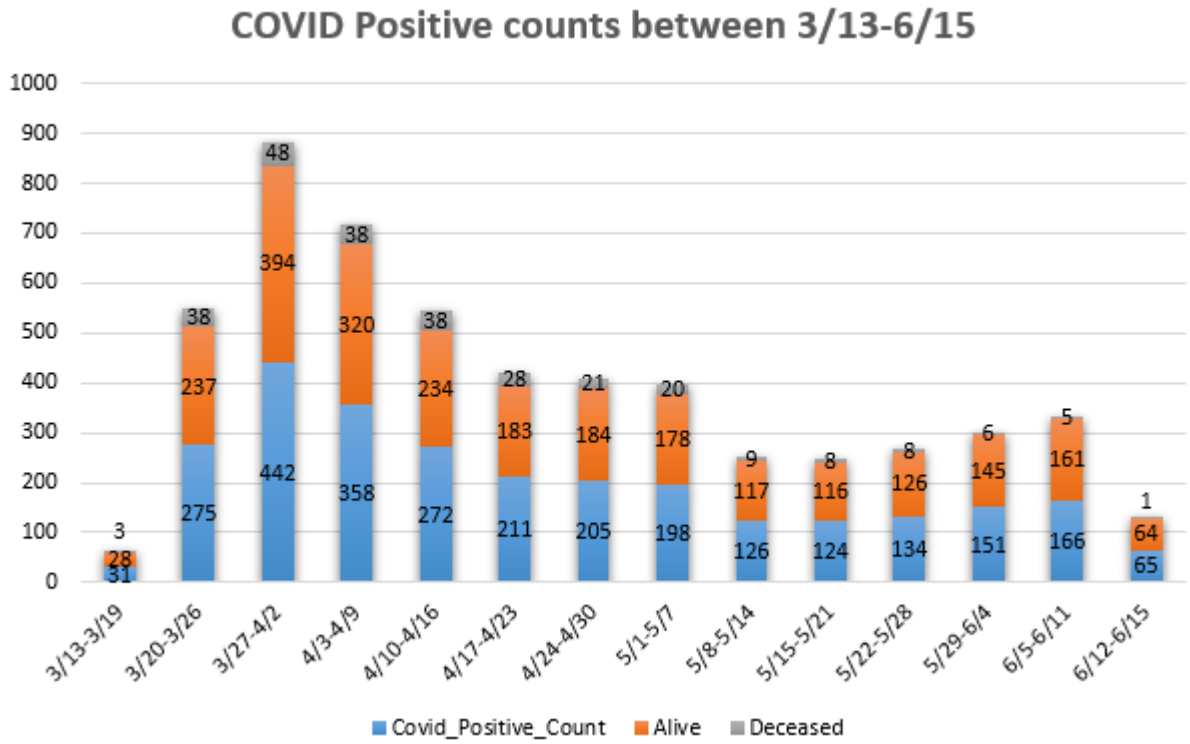
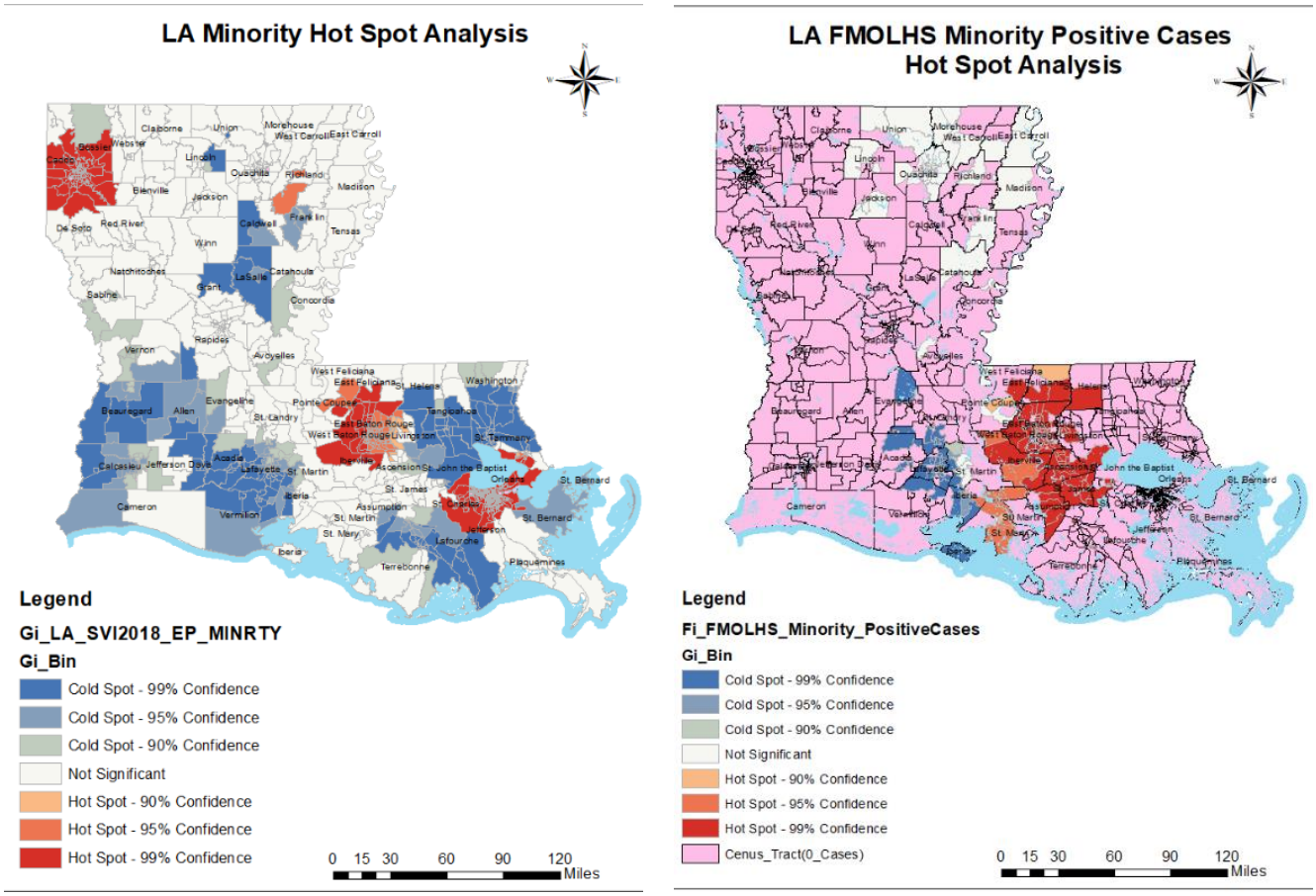


Figure 5.4 Weekly Mortality Rate of COVID Positive Patients

From Figure 5.4, we can see that there is an increase in the number of COVID positive patients and deaths from March 13th to April 2nd. Moreover, from April 3rd to May 14th there is a gradual decrease in the number of COVID positive cases and deaths. From May 15th to June 11th there is an increase in the number of COVID positive cases but, the number of deaths are decreased. Over all, the death rate decreased.

As part of the analysis and results discussed in Section 5.1, From Table 5.1, we can see that the CEVD, CKD, CVD, Diabetes, ESRD, and HTN underlying conditions, age and CPMI are weakly correlated with the mortality rate of COVID non-institutionalized positive patients. Moreover, the COVID test result ('Positive/Negative') is weakly correlated by the socioeconomic factors such as unemployment (EP_UNEMP), household composition factors single parent (EP_SNGPNT), minority status, and population who are doing service occupations. This implies population who work in jobs such as restaurants, plumber services etc. have more interaction with people. Additionally, FMOLHS patients who has obesity showed higher correlation on the COVID test result.

Based on Louisiana ACS data [88], 48.9% of LA population are male and 51.1% are females. Though the female population is more in LA, the death rates are high in FMOLHS COVID male patients. In general, the LA demographics based on race and ethnicity contributes 58.8% White, 32.0% Black or African American, 1.7% Asian, 0.7% Other, 5.0% Hispanic. In contrary, the FMOLHS majority of COVID positive percentage tested patients are higher in minority ethnicity population.



a) LA Population

b) FMOLHS COVID Minority Positive

Figure 5.5. Minority Comparison between a) LA Population b) FMOLHS COVID Minority Positive Cases Population Hot Spot Analysis

From Figure 5.5 a and b maps, we can see that few of the hot spots in census tracts of EBRP, West Baton Rouge, Livingston, Iberville of LA minority map are consistent with the FMOLHS minority population COVID positive tested patients. Also, from map b) we can see that there are other census tracts in Assumption, Ascension, St. Helen, St. James, Iberia parishes that shows the hot spots of minority FMOLHS populations who got COVID-19 disease.

From the statistics shown in Appendix Y and Appendix Z we can understand that the White ethnicity institutionalized population got COVID-19 more than other ethnicity (Asians, Other, Unknown, Hispanic). HTN, Diabetes, and CVD are the top 3 underlying conditions for

institutionalized patients. The institutionalized deceased positive patients average age is 76. Also, the Asians and Hispanics have 0% of death rates in non-institutionalized FMOLHS COVID positive patient population. From Appendix AA and Appendix AB we can see that the minority ethnicity non- institutionalized patients acquired COVID-19 more. The non-institutionalized deceased positive patients average age is 68. HTN, Diabetes, and Obesity are the top 3 underlying conditions for non-institutionalized patients.

Weak correlations are noticed for patient demographics Age, BMI, Race_Ethnicity, Smoking Status; Clinical data: LOS, Type; Underlying conditions: Obesity with respect to Covid test result. Due to small number of Covid positive tested records we did not observe strong correlations of features with respect to the Covid test result. The hot spot clusters of positive non-institutionalized Covid-19 patients are associated with high percentages of low per capita income; single parent population; minority status population; people living in mobile homes and crowded housing. Moreover, the positive patients spatial clusters showed higher occupation rate in service, sales, production, transportation and, material moving jobs. In the future research with the availability of more SDOH and clinical data the correlation results can change and these factors will be useful to the healthcare organization in developing a model that can predict the likelihood of testing positive for the Covid-19 virus and the outcomes of the disease(severity/mortality/hospitalization). Moreover, the good performance of the model can be useful for the organization to keep informed on the bad outcomes (severity/mortality/hospitalization) of the Covid-19 disease and improve the management of health/hospital scarce resources and disease in the event of contagious pandemic.

The hotspot analysis can be improved by including more patients and more hospitals. For example, not just the people going to OLOL facilities but including all the people who visit multiple hospitals not affiliated to FMOLHS. We can also perform the hotspot analysis based on underlying health conditions, social vulnerability index of people living in small geographical area. This can help us to know if the clusters of sick people living in same are being the major reason for the spread of the Covid-19 disease.

The key points of results discussed in Section 5.1 and Section 5.2 are:

- The institutionalized COVID positive tested patients are having high average age approximately around 70 when compared to non-institutionalized COVID positive tested.
- The male gender has higher mortality rate when compared to female gender.
- The minority FMOLHS COVID positive patient percentage is more when compared to LA demographics.
- The FMOLHS patients living in regions that has high percentages of socioeconomic factors such as poverty rate, unemployment rate, no high school diploma, and low per capita income; higher percentages of household composition factors such as elderly population, disability; higher percentages of minority status population and people living in crowded housing showed higher COVID death rates when compared to the counterparts.
- There were no deaths among Asian and Hispanic non-institutionalized positive patients.
- The patients who have occupations related to service, sales, production, transportation and, material moving have acquired COVID-19 at a higher rate than the general population.

Chapter 6. Conclusions and Future Research

GIS can play a major role in informing health care industry policy and decision making. By using ArcGIS and geocoding, FMOLHS patient data and spatial non-clinical data such as Social Determinants of Health (SDOH) factors may be integrated to enhance analysis and broaden the questions that may be investigated by the hospital system. Further, incorporating machine learning-based predictive analytics for the FMOLHS patient data can lead to the development of real-time decision-making tools for improving health-related outcomes. This predictive classification model developed in this research study can be used as an alternative to the existing prediction model in the FMOLHS Epic. This model can be used to predict the likelihood of the patient being an ED High Utilizer in the future.

In this research study, as part of ED High Utilizers predictive classification model development the ED high utilizers patient visits data is studied based on age, gender, year, financial class, and ethnicity to identify the visit pattern between 2015-2019. The correlation analysis is performed to identify the patient and census tract non-clinical (SDOH factors) data that are influencing the FMOLHS patients for the overutilization of ED for unnecessary/non-urgent reasons. In the next step of this research study, the FMOLHS COVID-19 tested patient data set is studied based on age, gender, institutionalized('Yes/No'), ethnicity, clinical, and non-clinical factors. Correlation analysis is performed to identify the factors that are responsible for the acquiring COVID-19 disease and mortality.

As part of this research study, ED high utilizer predictive classification model is developed. This model will be useful to the healthcare organization/social workers/patient advocates to follow up on the high-risk patients in order to keep them being an ED High utilizer. For example, these results can be used to make a phone call to the high-risk patients a week following ED visit and make sure that they are taking medicine, eating right food, doing exercise, and properly monitoring their health. More than 70% of OLOL ED high utilizers are covered by Medicaid and Medicare. In general, if the ED high utilizers covered by Medicaid, Medicare, and other payers are readmitted to ED within 30 days then the hospital organization have to pay penalties charged by Center for Medicare & Medicaid Services (CMS) and other payers. Hence these results can be used by community paramedics to educate people on healthier lifestyles that can reduce the risk of developing severe illness and provide primary care services in order to reduce non-urgent ED visits when the patient health problem is not critical. Having good predictor results will help the healthcare organization to reduce financial penalties by reducing the number of ED high utilizers and associated over utilization of hospital scarce resources.

As part of COVID study, correlation analysis and spatial analysis is done to identify the features associated with the risk of acquiring/spreading Covid-19. This analysis results will be useful for the hospital organization to know how different neighborhood characteristics impact the spread of Covid-19 and can develop a strategic response throughout the different stages of pandemic. For example, managing the patient flow, staffing, and hospital resources (PPE (Personal Protective Equipment), supplies, ventilators) during the pandemic.

The techniques/methodologies used in the above summarized analysis can be used by healthcare providers and research institutions to study, analyze, predict, and uncover the hidden patterns in different types of healthcare data available with simple tweaks. Implementing of the

latest artificial intelligence tools and statistical techniques like in this study opens up new ways to understand and investigate huge data which otherwise cannot be anticipated by human intelligence.

The scope of this research study is LA but we have included only FMOLHS affiliated patient data across LA. For the machine learning model, we have not validated the model on "Live" patient data in the operational environment. Moreover, the underlying health conditions are not included in the model. Due to unavailability of latest spatial non-clinical data the COVID-19 retrospective study had integrated the 2018 non-clinical data to the clinical data. Though there are these limitations, this research study has successfully generated outcomes that would be useful for developing the predictive model of COVID-19 disease. By using ED overutilization data analysis results, we could compare the performance of existing prediction model in the FMOLHS Epic with the predictive classification model developed in this research study. Additionally, we could explore on developing an interactive web application that will generate social prescriptions based on the patient's physical address. Moreover, with the help of ED High Utilizers data analysis results, we could perform a research study on identifying patterns or uncover the hidden facts in the FMOLHS ED revisits within 30 days.

In this research study, with the help of latest technology I was able to learn how to discover the hidden facts in different types of healthcare data. Additionally, I was able to learn the importance of data-driven analytical solutions in improving the existing process efficiency of a healthcare organization. By doing thesis as part of my master's degree I was able to improve my writing and presentation skills. Moreover, I was able to collaborate with new people and communicate with them. This helped me to learn/improve my project management, team work, and communication skills.

Appendix A. IRB Approval Letter from OLOL RMC

OUR LADY OF THE LAKE
REGIONAL MEDICAL CENTER
Franciscan Missionaries of Our Lady Health System

January 31, 2018

Academic Affairs
Office of Research

Diana Hamer, PhD
5246 Brittany Dr., Office #410
Baton Rouge, LA 70808

RE: Integrating Geospatial Analytics into Healthcare – Delivering Community Information at the Point-of-Care

Dear Dr. Hamer,

Proposed studies that will utilize Our Lady of the Lake Regional Medical Center (OLOL) resources will be reviewed and endorsed by the OLOL Office of Research.

In accordance with the policies of the OLOL Office of Research, it was confirmed that as the Principal Investigator for this study, you are not listed on the FDA's Debarment List, or the Clinical Investigators-Disqualification Proceedings List.

The above-referenced study has been reviewed, and it is my pleasure to inform you that your study was granted endorsement from the OLOL Office of Research. Additionally, I confirm that all required study personnel documents are on file in the OLOL Office of Research.

You may not commence study activity at our site until the OLOL Office of Research receives the approval letter from the Louisiana State University Health Sciences Center-New Orleans Institutional Review Board (LSUHSC-NO IRB).

As the Principal Investigator, it is your responsibility to assure that all study-related activity strictly adheres to local, state and federal regulations for the oversight of human subject research, as well as the institutional policies.

Please file this notice with your study files and include a copy of this correspondence with your packet for submission to LSUHSC-NO IRB. Please submit a copy of your study findings to the OLOL Office of Research upon completion of your project.

Sincerely,

Jtomkins

Jill Tomkins, BA, CCRP
Regulatory Manager

7777 HENNESSY BOULEVARD, SUITE 6004, BATON ROUGE, LA 70808 • PHONE (225) 765-7659 • FAX (225) 765-4345

Appendix B. IRB Approval Letters from FMOLU

FRANCISCAN
MISSIONARIES OF OUR LADY
UNIVERSITY

Institutional Review Board (IRB)

Date: October 04, 2018

Study Number: 2018-151
Study Title: *Integrating Geospatial Analytics into Healthcare – delivering community information at the point-of-care*
Primary Investigator: Diana Hamer, PhD
Secondary Investigators: Deekshith Mandala, Tonya Jagneaux, MD
Primary Reviewer: Lindsay Mullins, PhD
Secondary Reviewer: Keeley Harmon, PhD
Approval Designation: Exempt
Approval Date: October 04, 2018
Expiration Date: October 03, 2019

Dear Dr. Hamer,

I am pleased to inform you that Lindsay Mullins, PhD and Keeley Harmon, PhD of Franciscan Missionaries of Our Lady University Institutional Review Board have reviewed and approved your proposed study entitled *Integrating Geospatial Analytics into Healthcare – delivering community information at the point-of-care* conducted by Diana Hamer, PhD, Deekshith Mandala and Tonya Jagneaux, MD.

Please be aware that this approval is only valid for one year. If your research extends past that time, you will need to submit a Re-approval application form no later than two weeks before the end of the approval period.

Thank you for your submission and I would like to wish you success with your study.

Best regards,

Dr. Michael T. Dreznick,
Associate Professor
Franciscan Missionaries of Our Lady University and IRB Chair

FRANCISCAN
MISSIONARIES OF OUR LADY
UNIVERSITY

Institutional Review Board (IRB)

Date: January 29, 2020

Primary Investigator: Diana Hamer, PhD
Approval Designation: Exempt

Dear Dr. Hamer,

Thank you for submitting your modification request regarding study number 2018-151 titled *Integrating Geospatial Analytics into Healthcare – delivering community information at the point-of-care*. Approval is granted for the use of this updated material.

Add sub-investigator – RamyaKrishna Tummala, BS

Thank you for your update I would like to wish you continued success with your study.

Best regards,

Dr. Michael T. Dreznick,
Associate Professor
Franciscan Missionaries of Our Lady University and IRB Chair

FRANCISCAN
MISSIONARIES OF OUR LADY
UNIVERSITY

Institutional Review Board (IRB)

Date: September 03, 2019

Primary Investigator: Diana Hamer, PhD
Approval Designation: Exempt

Dear Dr. Hamer,

Thank you for submitting your continuation request regarding study number 2018-151 titled *Integrating Geospatial Analytics into Healthcare – delivering community information at the point-of-care*. Approval is granted with an expiration date of September 02, 2020.

Please be aware that this approval is only valid for one year. If your research extends past that time, you will need to submit a Re-approval application form no later than two weeks before the end of the approval period.

Thank you for your submission and I would like to wish you continued success with your study.

Best regards,

Dr. Michael T. Dreznick,
Associate Professor
Franciscan Missionaries of Our Lady University and IRB Chair

FRANCISCAN
MISSIONARIES OF OUR LADY
UNIVERSITY

Institutional Review Board (IRB)

Date: August 11, 2020

Primary Investigator: Diana Hamer, PhD
Approval Designation: Exempt

Dear Dr. Hamer,

Thank you for submitting your continuation request regarding study number **2018-151** titled *Integrating Geospatial Analytics into Healthcare – delivering community information at the point-of-care*. Approval is granted with an expiration date of August 10, 2021.

Please be aware that this approval is only valid for one year. If your research extends past that time, you will need to submit a Re-approval application form no later than two weeks before the end of the approval period.

Thank you for your submission and I would like to wish you continued success with your study.

Best regards,

Dr. Michael T. Dreznick,
Associate Professor
Franciscan Missionaries of Our Lady University and IRB Chair

Appendix C. IRB Approval Letters from LSUHSC

EXPEDITED APPROVAL
LOUISIANA STATE UNIVERSITY HEALTH SCIENCES CENTER
(Assurance Number FWA00002762)
IRB Registration Number 00000177

FROM: LSUHSC-NO Institutional Review Board
TO: Joseph Moerschbaecher, Ph.D.
Vice Chancellor for Academic Affairs
RE: IRB Application By: **Diana Hamer, PhD**
Department of Internal Medicine

Entitled: IRB # 10102: Integrating Geospatial Analytics into Healthcare - Delivering Community Information at the Point-of-Care

This is to document review and approval of the above-referenced research protocol. In the judgment of this Board, the procedures delineated in said application conform to the pertinent DHHS and FDA rules and regulations regarding use of human subjects. This procedure is authorized by 45CFR46.110 and 21CFR56.110 as published in the Federal Register November 9, 1998. Records regarding action of the Board, referable to said project, are on file in the Office of the Chairman. This study is expedited under 46.110 category **#5 of** 45CFR Part 46.

THE INVESTIGATOR agrees to report to the Committee any emergent problems, serious adverse reactions, or procedural changes that may affect the status of the investigation, and that no such changes will be made without Board approval, except where necessary to eliminate apparent immediate hazards to the subject. The investigator also agrees to periodic review of this project by the Board at intervals appropriate to the degree of risk to assure that the new project is being conducted in compliance with the Board's understanding and recommendation, and this interval will not exceed one year.

PLEASE NOTE:

1. Any advertisement to recruit subjects for this study must be approved by the IRB prior to posting, publication and/or distribution.
2. Other institutional approvals may be required before the study can be initiated.
3. Written notification (at the time this study is completed/ canceled) must be sent to the Office of the Chair.
4. Waiver of Informed Consent granted.
5. Waiver of HIPAA Authorization granted

Approval Period: 6/28/2018 - 6/27/2019

Diana Hamer, PhD DATE: 6/28/2018

Jawed Alam, PHD, MBA, Chairman DATE: 6/28/2018
Edward S. Peters, DMD, SM, ScD

APPROVED - LSUHSC IRB
February 11, 2020

LSU Health NEW ORLEANS

LSUHSC-NO Human Research Protection Program
Institutional Review Board
Assurance Number: FWA00002762

DATE: February 11, 2020

TO: Diana Hamer, PhD
Principal Investigator

FROM: LSUHSC-NO Institutional Review Board

RE: IRB #: 10102
Protocol Title: Integrating Geospatial Analytics into Healthcare - Delivering Community Information at the Point-of-Care
Submission Type: Modification Request – Study Personnel

DETERMINATION: **Approved**
REVIEW TYPE: **Administrative**

Thank you for your recent modification request, which requested the following:

Change in Personnel (Other than PI) ONLY

Based on your submission, the current study team is as follows:

- Diana Hamer, PhD - Principal Investigator
- Alicia Chautin, MD - Research Assistant
- Brogan McNease - Research Assistant
- Christopher Thomas, MD - Research Assistant
- Deekshith Mandala, MS - Research Assistant
- Hollis O'Neal Jr., MD, MSCI - Co-Investigator
- Jill Tomkins - Site Manager
- Joel Mosley, MD - Research Assistant
- John Marston, MD - Research Assistant
- Kelsey Gore - Student Investigator
- Kody Bliss - Research Assistant
- Mandi Musso, PhD - Co-Investigator
- Michael Baranowski, MD - Other Study Personnel
- Quan Shi, MD - Other Study Personnel
- Roshan Patel - Research Assistant
- Taylor Hillburn - Research Assistant
- Tonya Jagneaux, MD - Co-Investigator
- Tummala Ramyakrishna, BS - Other Study Personnel

Appendix D. IRB Approval Letter from LSU

ACTION ON PROTOCOL APPROVAL REQUEST

Institutional Review Board
Dr. Dennis Landin, Chair
130 David Boyd Hall
Baton Rouge, LA 70803
P: 225.578.8692
F: 225.578.5983
irb@lsu.edu
lsu.edu/research

TO: Gerald Knapp
Industrial Engineering

FROM: Dennis Landin
Kinesiology

DATE: October 24, 2019

RE: IRB# 4299

TITLE: Integrating Geospatial Analytics into Healthcare - delivering community information at the point-of-care

New Protocol/Modification/Continuation: New Protocol

Review type: Full ___ Expedited _X_ **Review date:** 10/23/2019

Risk Factor: Minimal ___X___ Uncertain _____ Greater Than Minimal_____

Approved_____X_____ **Disapproved**_____

Approval Date: 10/23/2019 **Approval Expiration Date:** 10/22/2020

Re-review frequency: (annual unless otherwise stated)

Number of subjects approved: 400,000

LSU Proposal Number (if applicable)**:**

By: Dennis Landin, Chairman _____

PRINCIPAL INVESTIGATOR: PLEASE READ THE FOLLOWING –
Continuing approval is CONDITIONAL on:

1. Adherence to the approved protocol, familiarity with, and adherence to the ethical standards of the Belmont Report, and LSU's Assurance of Compliance with DHHS regulations for the protection of human subjects*
2. Prior approval of a change in protocol, including revision of the consent documents or an increase in the number of subjects over that approved.
3. Obtaining renewed approval (or submittal of a termination report), prior to the approval expiration date, upon request by the IRB office (irrespective of when the project actually begins); notification of project termination.
4. Retention of documentation of informed consent and study records for at least 3 years after the study ends.
5. Continuing attention to the physical and psychological well-being and informed consent of the individual participants, including notification of new information that might affect consent.
6. A prompt report to the IRB of any adverse event affecting a participant potentially arising from the study.
7. Notification of the IRB of a serious compliance failure.
8. **SPECIAL NOTE: When emailing more than one recipient, make sure you use bcc.**

**All investigators and support staff have access to copies of the Belmont Report, LSU's Assurance with DHHS, DHHS (45 CFR 46) and FDA regulations governing use of human subjects, and other relevant documents in print in this office or on our World Wide Web site at http://www.lsu.edu/irb*

Appendix E. Waiver of HIPAA Authorization from LSUHSC

LOUISIANA STATE UNIVERSITY HEALTH SCIENCES CENTER
(Assurance Number FWA00002762)

Waiver of HIPAA Authorization

A waiver of authorization has been granted to Diana Hamer, PhD

IRB# 10102: Integrating Geospatial Analytics into Healthcare - Delivering Community Information at the Point-of-Care

by the LSUHSC- New Orleans Institutional Review Board, Reg.#00000177. This study meets the waiver criteria described in 45 CFR 164.512 (i) (2) (ii).

The request for this waiver was reviewed using the EXPEDITED procedure approved by the LSUHSC Institutional Review Board, and as required by 45 CFR 164.512 (i) (2) (iv).

This waiver pertains to Protected Health Information (PHI) as described in the study protocol.

With a waiver of authorization, investigators must keep a record of the subjects from whom Protected Health Information has been gathered and to whom this PHI is disclosed. This information must be provided to the subject upon request.

Approval Date: 6/28/2018

Approved By LSUHSC IRB Chairman or Vice-Chair

Acknowledgement of Principal Investigator:
Diana Hamer, PhD

Date: 6/28/2018

Appendix F. ED Discharge patients Categorical Variables and Frequencies

Categorical Variable	Category Values		
	Label	Frequency	Percent
Admit	ED Discharge (0)	391,243	
Year	2015	72,043	18.4
	2016	73,351	18.7
	2017	79,793	20.4
	2018	74,894	19.1
	2019	91,162	23.3
Fin_class	Medicaid (3)	171,569	43.9
	Private Insurance (1)	107,143	27.4
	Self Pay (4)	59,544	15.2
	Medicare (2)	52,981	13.5
	Other (5)	6	0.0015
Age	Less than 0 (0)	1	0.0002
	0-10 (1)	88,797	22.7
	11-20 (2)	51,239	13.1
	21-30 (3)	65,521	16.7
	31-40 (4)	54,957	14.0
	41-50 (5)	41,573	10.6
	51-60 (6)	39,350	10.1
	61-70 (7)	26,201	6.7
	71-80 (8)	14,657	3.7
	81-90 (9)	7,292	1.9
	91-100 (10)	1,605	0.4
	101-110 (11)	27	0.007
	Unknown (12)	23	0.006
Race	Black (2)	182,462	46.6
	White (1)	176,045	45.0
	Hispanic (3)	14,610	3.7
	Asian (4)	5,859	1.5
	Other or Unknown (5)	10,969	2.8
	Missing Data (999)	1,298	0.3
Gender	Female (2)	207,122	52.9
	Male (1)	184,023	47.0
	Missing Data (999)	98	0.025
RUCA	Metropolitan area core (1)	328,644	84
	Metropolitan area high commuting (2)	51,333	13.1
	Micropolitan area core (4)	4,665	1.2
	Rural areas (10)	3,917	1.0
	Small town core (7)	863	0.2
	Metropolitan area low commuting (3)	760	0.2
	Micropolitan area high commuting (2)	355	0.1
	Small town high commuting (8)	266	0.1
	Small town low commuting (9)	206	0.1
	Micropolitan area low commuting (6)	186	1.0
	Not Coded (99)	48	0.012

Appendix G. Categorical Variables and Frequencies of COVID Positive Tested Institutionalized Patients

Categorical Variable	Label	Frequency	Percent (%)
Age_Cat	0-10	0	0
	11-20	0	0
	21-30	2	0.4
	31-40	7	1.3
	41-50	21	3.8
	51-60	58	10.5
	61-70	137	24.7
	71-80	140	25.3
	81-90	142	25.6
	91-100	44	7.9
	101-110	3	0.5
Type	Admit	281	50.7
	Outpatient /UC	252	45.5
	ED	21	3.8
Living_Status	Alive	442	79.8
	Deceased	112	20.2
Sex	Female	282	50.9
	Male	272	49.1
Institution_group	Nursing Home	494	89.2
	Prison	52	9.1
	Mental Health / Rehab	8	1.4
Smoking_Status	Never	271	48.9
	Former	173	31.2
	Current	73	13.2
	Unknown	37	6.7
Race_Ethnicity	White or Causian	257	46.4
	Black or African American	242	43.7
	Hispanic	7	1.3
	Other	7	1.3
	Asian	3	0.5
	Unknown	38	6.9
Discharge_Disposition	Nursing Home	107	19.3
	Deceased	88	15.9
	Law enforcement	46	8.3
	Another acute care facility	35	6.3
	Hospice	17	3.1
	Mental Health/ Rehab	6	1.1
	Home	3	0.5
	Missing	252	45.7

(Table Cont'd)

Categorical Variable	Label	Frequency	Percent
Auto immune	Yes	21	3.8
	No	533	96.2
CEVD	Yes	148	26.7
	No	406	73.3
CKD	Yes	169	30.5
	No	385	69.5
CLD	Yes	13	2.3
	No	541	97.7
CVD	Yes	202	36.5
	No	352	63.5
Diabetes	Yes	251	45.3
	No	303	54.7
ESRD	Yes	20	3.6
	No	534	96.4
Hep B	Yes	6	1.1
	No	548	98.9
Hep C	Yes	18	3.2
	No	536	96.8
HIV	Yes	5	0.9
	No	549	99.1
HTN	Yes	479	86.5
	No	75	13.5
Immunocompromised	Yes	1	0.2
	No	553	99.8
Obesity	Yes	106	19.1
	No	448	80.9
Overweight	Yes	551	0.5
	No	3	99.5
PVD	Yes	49	8.8
	No	505	91.2
Respiratory	Yes	112	20.2
	No	442	79.8
RUCA	Metropolitan area core	428	77.3
	Metropolitan area high commuting	69	12.5
	Rural areas	28	5.1
	Micropolitan area core	22	4.0
	Micropolitan area high commuting	3	0.5
	Small town core	2	0.4
	Metropolitan area low commuting	0	0
	Micropolitan area low commuting	0	0
	Small town high commuting	0	0
	Small town low commuting	0	0
	Not Coded	0	0
	Missing Data	2	0.4

Appendix H. Categorical Variables and Frequencies of COVID Positive Tested Non-Institutionalized Patients

Categorical variable	Label	Frequency	Percent
Age_Cat	0-10	0	0
	11-20	78	3.5
	21-30	308	14.0
	31-40	316	14.3
	41-50	368	16.7
	51-60	439	19.9
	61-70	369	16.7
	71-80	198	9.0
	81-90	97	4.4
	91-100	31	1.4
	101-110	0	0
Type	Outpatient / UC	1,019	46.2
	Admit	620	28.1
	ED	565	25.6
Living_Status	Alive	2,041	92.6
	Deceased	163	7.4
Sex	Female	1,302	59.1
	Male	902	40.9
Smoking_Status	Never	1465	66.5
	Former	394	17.9
	Current	138	6.3
	Unknown	207	9.4
Race_Ethnicity	Black or African American	1,311	59.5
	White or Causian	653	29.6
	Hispanic	93	4.2
	Other	28	1.3
	Asian	13	0.6
	Unknown	106	4.8
Discharge_Disposition	Home	905	41.1
	Deceased	151	6.9
	Another acute care facility	72	3.3
	Nursing Home	38	1.7
	Hospice	28	1.3
	Mental Health/ Rehab	25	1.1
	Law enforcement	0	0
	Missing	985	44.7
Auto immune	Yes	58	2.6
	No	2,146	97.4

(Table Cont'd)

Categorical Variable	Label	Frequency	Percent
CEVD	Yes	132	6.0
	No	2,072	94.0
CKD	Yes	236	10.7
	No	1,968	89.3
CLD	Yes	36	1.6
	No	2,168	98.4
CVD	Yes	283	12.8
	No	1,921	87.2
Diabetes	Yes	624	28.3
	No	1,580	71.7
ESRD	Yes	52	2.4
	No	2,152	97.6
Hep B	Yes	17	0.8
	No	2,187	99.2
Hep C	Yes	23	1.0
	No	2,181	99.0
HIV	Yes	27	1.2
	No	2,177	98.8
HTN	Yes	1,075	48.8
	No	1,129	51.2
Immunocompromised	Yes	6	0.3
	No	2,198	99.7
Obesity	Yes	583	26.5
	No	1,621	73.5
Overweight	Yes	22	1.0
	No	2,182	99.0
PVD	Yes	53	2.4
	No	2,151	97.6
Respiratory	Yes	224	10.2
	No	1,980	89.8
RUCA	Metropolitan area core	1,516	68.8
	Metropolitan area high commuting	257	11.7
	Micropolitan area core	144	6.5
	Rural areas	32	1.5
	Small town core	30	1.4
	Micropolitan area high commuting	26	1.2
	Micropolitan area low commuting	8	0.4
	Small town high commuting	4	0.2
	Metropolitan area low commuting	0	0
	Small town low commuting	0	0
	Not Coded	0	0
	Missing Data	187	8.5

Appendix I. Spearman Correlation Scores between independent variables of ED High Utilizers Data

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	Year	LOS_Yr	Age	Admit	Gender
Year	***	0.15	0.15	0.15	0.01
LOS_Yr	0.15	***	0.35	0.99	0.01
Age	0.15	0.35	***	0.33	0.07
Admit	0.15	0.99	0.33	***	0.01
Gender	0.01	0.01	0.07	0.01	***
UC_Visits	0.08	-0.09	-0.05	-0.09	0.02
PC_Visits	0.15	0.04	-0.14	0.04	0.00
ED_Unneces_Yr	-0.34	-0.68	-0.34	-0.68	-0.01
ED_HighUtilizers	-0.40	-0.47	-0.34	-0.47	-0.01
E_TOTPOP	0.08	-0.01	-0.04	-0.01	0.02
EP_POV	-0.16	-0.13	-0.09	-0.13	0.00
EP_UNEMP	-0.16	-0.12	-0.08	-0.12	-0.01
EP_PCI	0.21	0.16	0.13	0.16	0.00
EP_NOHSDP	-0.10	-0.14	-0.07	-0.15	0.00
EP_AGE65	0.12	0.14	0.16	0.13	0.00
EP_AGE17	-0.01	-0.07	-0.09	-0.07	0.02
EP_DISABL	-0.06	-0.06	0.01	-0.06	-0.01
EP_SNGPNT	-0.12	-0.11	-0.10	-0.11	0.00
EP_MINRTY	-0.01	-0.03	-0.03	-0.03	-0.02
EP_LIMENG	0.11	0.02	0.01	0.02	0.00
EP_MUNIT	-0.10	-0.03	-0.03	-0.03	0.00
EP_MOBILE	0.07	-0.03	-0.01	-0.03	0.02
EP_CROWD	-0.10	-0.11	-0.11	-0.11	0.01
EP_NOVEH	-0.05	-0.09	-0.02	-0.09	0.00
EP_GROUPQ	-0.07	-0.06	-0.01	-0.06	-0.02
RPL_THEME1	-0.12	-0.15	-0.10	-0.15	0.00
RPL_THEME2	-0.07	-0.05	-0.03	-0.06	0.00
RPL_THEME3	0.07	0.00	-0.01	-0.01	-0.01
RPL_THEME4	-0.10	-0.13	-0.08	-0.14	0.00
RPL_THEMES	-0.11	-0.13	-0.08	-0.14	0.00
EP_UNINSUR	-0.32	-0.16	-0.11	-0.16	0.01

(Table Cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	UC_Visits	PC_Visits	E_TOTPOP	EP_POV
Year	0.08	0.15	0.08	-0.16
LOS_Yr	-0.09	0.04	-0.01	-0.13
Age	-0.05	-0.14	-0.04	-0.09
Admit	-0.09	0.04	-0.01	-0.13
Gender	0.02	0.00	0.02	0.00
UC_Visits	***	0.01	-0.17	0.22
PC_Visits	0.01	***	-0.02	-0.03
ED_Unneces_Yr	0.15	-0.12	-0.07	0.23
ED_HighUtilizers	0.16	-0.14	-0.09	0.25
E_TOTPOP	-0.17	-0.02	***	-0.49
EP_POV	0.22	-0.03	-0.49	***
EP_UNEMP	0.19	-0.04	-0.33	0.63
EP_PCI	-0.22	0.02	0.40	-0.78
EP_NOHSDP	0.13	-0.02	-0.34	0.65
EP_AGE65	-0.02	0.05	-0.37	-0.13
EP_AGE17	0.04	-0.02	0.41	-0.08
EP_DISABL	0.20	0.00	-0.54	0.54
EP_SNGPNT	0.17	-0.02	-0.07	0.47
EP_MINRTY	0.28	-0.01	-0.47	0.56
EP_LIMENG	-0.01	0.00	0.02	-0.05
EP_MUNIT	0.11	0.00	-0.20	0.18
EP_MOBILE	-0.17	0.00	0.32	-0.17
EP_CROWD	0.11	0.00	-0.04	0.41
EP_NOVEH	0.21	-0.02	-0.56	0.73
EP_GROUPQ	0.08	-0.05	-0.19	0.28
RPL_THEME1	0.22	-0.02	-0.46	0.88
RPL_THEME2	0.18	0.01	-0.32	0.46
RPL_THEME3	0.17	-0.01	-0.29	0.32
RPL_THEME4	0.12	-0.03	-0.22	0.54
RPL_THEMES	0.23	-0.02	-0.47	0.79
EP_UNINSUR	0.11	-0.04	-0.39	0.63

(Table Cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	EP_UNEMP	EP_PCI	EP_NOHSDP	EP_AGE65
Year	-0.16	0.21	-0.10	0.12
LOS_Yr	-0.12	0.16	-0.14	0.14
Age	-0.08	0.13	-0.07	0.16
Admit	-0.12	0.16	-0.15	0.13
Gender	-0.01	0.00	0.00	0.00
UC_Visits	0.19	-0.22	0.13	-0.02
PC_Visits	-0.04	0.02	-0.02	0.05
ED_Unneces_Yr	0.21	-0.27	0.21	-0.17
ED_HighUtilizers	0.22	-0.30	0.22	-0.17
E_TOTPOP	-0.33	0.40	-0.34	-0.37
EP_POV	0.63	-0.78	0.65	-0.13
EP_UNEMP	***	-0.63	0.54	-0.13
EP_PCI	-0.63	***	-0.75	0.20
EP_NOHSDP	0.54	-0.75	***	-0.05
EP_AGE65	-0.13	0.20	-0.05	***
EP_AGE17	0.06	-0.13	0.06	-0.56
EP_DISABL	0.48	-0.56	0.61	0.29
EP_SNGPNT	0.35	-0.55	0.39	-0.32
EP_MINRTY	0.51	-0.54	0.26	0.00
EP_LIMENG	-0.15	0.08	-0.12	0.00
EP_MUNIT	0.04	-0.06	-0.14	0.02
EP_MOBILE	-0.13	0.02	0.29	-0.11
EP_CROWD	0.30	-0.46	0.45	-0.31
EP_NOVEH	0.53	-0.60	0.63	0.09
EP_GROUPQ	0.19	-0.30	0.22	0.05
RPL_THEME1	0.80	-0.90	0.84	-0.13
RPL_THEME2	0.40	-0.56	0.53	0.19
RPL_THEME3	0.22	-0.30	0.09	0.02
RPL_THEME4	0.35	-0.55	0.55	-0.12
RPL_THEMES	0.64	-0.84	0.75	-0.02
EP_UNINSUR	0.43	-0.69	0.62	-0.13

(Table Cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	EP_AGE17	EP_DISABL	EP_SNGPNT	EP_MINRTY
Year	-0.01	-0.06	-0.12	-0.01
LOS_Yr	-0.07	-0.06	-0.11	-0.03
Age	-0.09	0.01	-0.10	-0.03
Admit	-0.07	-0.06	-0.11	-0.03
Gender	0.02	-0.01	0.00	-0.02
UC_Visits	0.04	0.20	0.17	0.28
PC_Visits	-0.02	0.00	-0.02	-0.01
ED_Unneces_Yr	0.06	0.12	0.16	0.10
ED_HighUtilizers	0.05	0.13	0.18	0.12
E_TOTPOP	0.41	-0.54	-0.07	-0.47
EP_POV	-0.08	0.54	0.47	0.56
EP_UNEMP	0.06	0.48	0.35	0.51
EP_PCI	-0.13	-0.56	-0.55	-0.54
EP_NOHSDP	0.06	0.61	0.39	0.26
EP_AGE65	-0.56	0.29	-0.32	0.00
EP_AGE17	***	-0.12	0.46	-0.06
EP_DISABL	-0.12	***	0.26	0.42
EP_SNGPNT	0.46	0.26	***	0.43
EP_MINRTY	-0.06	0.42	0.43	***
EP_LIMENG	-0.07	-0.10	0.00	0.22
EP_MUNIT	-0.25	-0.01	0.05	0.34
EP_MOBILE	0.29	-0.01	0.01	-0.56
EP_CROWD	0.33	0.22	0.44	0.22
EP_NOVEH	-0.15	0.59	0.29	0.52
EP_GROUPQ	-0.25	0.16	0.12	0.24
RPL_THEME1	0.02	0.63	0.48	0.54
RPL_THEME2	0.30	0.72	0.71	0.39
RPL_THEME3	-0.10	0.21	0.25	0.76
RPL_THEME4	-0.01	0.34	0.37	0.30
RPL_THEMES	0.04	0.65	0.60	0.64
EP_UNINSUR	-0.04	0.41	0.42	0.40

(Table Cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	EP_LIMENG	EP_MUNIT	EP_MOBILE	EP_CROWD
Year	0.11	-0.10	0.07	-0.10
LOS_Yr	0.02	-0.03	-0.03	-0.11
Age	0.01	-0.03	-0.01	-0.11
Admit	0.02	-0.03	-0.03	-0.11
Gender	0.00	0.00	0.02	0.01
UC_Visits	-0.01	0.11	-0.17	0.11
PC_Visits	0.00	0.00	0.00	0.00
ED_Unneces_Yr	-0.05	0.10	-0.03	0.16
ED_HighUtilizers	-0.05	0.12	-0.06	0.18
E_TOTPOP	0.02	-0.20	0.32	-0.04
EP_POV	-0.05	0.18	-0.17	0.41
EP_UNEMP	-0.15	0.04	-0.13	0.30
EP_PCI	0.08	-0.06	0.02	-0.46
EP_NOHSDP	-0.12	-0.14	0.29	0.45
EP_AGE65	0.00	0.02	-0.11	-0.31
EP_AGE17	-0.07	-0.25	0.29	0.33
EP_DISABL	-0.10	-0.01	-0.01	0.22
EP_SNGPNT	0.00	0.05	0.01	0.44
EP_MINRTY	0.22	0.34	-0.56	0.22
EP_LIMENG	***	0.36	-0.20	0.01
EP_MUNIT	0.36	***	-0.61	0.02
EP_MOBILE	-0.20	-0.61	***	0.11
EP_CROWD	0.01	0.02	0.11	***
EP_NOVEH	-0.03	0.23	-0.14	0.28
EP_GROUPQ	0.02	0.27	-0.19	0.06
RPL_THEME1	-0.11	0.03	-0.01	0.45
RPL_THEME2	-0.05	-0.08	0.08	0.34
RPL_THEME3	0.78	0.43	-0.47	0.15
RPL_THEME4	0.07	0.41	0.04	0.61
RPL_THEMES	0.14	0.24	-0.07	0.55
EP_UNINSUR	0.08	0.22	-0.01	0.34

(Table Cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	EP_NOVEH	EP_GROUPQ	RPL_THEME1	RPL_THEME2
Year	-0.05	-0.07	-0.12	-0.07
LOS_Yr	-0.09	-0.06	-0.15	-0.05
Age	-0.02	-0.01	-0.10	-0.03
Admit	-0.09	-0.06	-0.15	-0.06
UC_Visits	0.00	-0.02	0.00	0.00
Gender	0.21	0.08	0.22	0.18
PC_Visits	-0.02	-0.05	-0.02	0.01
ED_Unneces_Yr	0.17	0.12	0.25	0.10
ED_HighUtilizers	0.18	0.13	0.26	0.11
E_TOTPOP	-0.56	-0.19	-0.46	-0.32
EP_POV	0.73	0.28	0.88	0.46
EP_UNEMP	0.53	0.19	0.80	0.40
EP_PCI	-0.60	-0.30	-0.90	-0.56
EP_NOHSDP	0.63	0.22	0.84	0.53
EP_AGE65	0.09	0.05	-0.13	0.19
EP_AGE17	-0.15	-0.25	0.02	0.30
EP_DISABL	0.59	0.16	0.63	0.72
EP_SNGPNT	0.29	0.12	0.48	0.71
EP_MINRTY	0.52	0.24	0.54	0.39
EP_LIMENG	-0.03	0.02	-0.11	-0.05
EP_MUNIT	0.23	0.27	0.03	-0.08
EP_MOBILE	-0.14	-0.19	-0.01	0.08
EP_CROWD	0.28	0.06	0.45	0.34
EP_NOVEH	***	0.26	0.71	0.42
EP_GROUPQ	0.26	***	0.28	0.06
RPL_THEME1	0.71	0.28	***	0.54
RPL_THEME2	0.42	0.06	0.54	***
RPL_THEME3	0.31	0.16	0.28	0.20
RPL_THEME4	0.60	0.64	0.56	0.31
RPL_THEMES	0.74	0.44	0.87	0.67
EP_UNINSUR	0.58	0.32	0.66	0.39

(Table Cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	RPL_THEME3	RPL_THEME4	RPL_THEMES	EP_UNINSUR
Year	0.07	-0.10	-0.11	-0.32
LOS_Yr	0.00	-0.13	-0.13	-0.16
Age	-0.01	-0.08	-0.08	-0.11
Admit	-0.01	-0.14	-0.14	-0.16
Gender	-0.01	0.00	0.00	0.01
UC_Visits	0.17	0.12	0.23	0.11
PC_Visits	-0.01	-0.03	-0.02	-0.04
ED_Unneces_Yr	0.03	0.22	0.23	0.27
ED_HighUtilizers	0.04	0.23	0.25	0.29
E_TOTPOP	-0.29	-0.22	-0.47	-0.39
EP_POV	0.32	0.54	0.79	0.63
EP_UNEMP	0.22	0.35	0.64	0.43
EP_PCI	-0.30	-0.55	-0.84	-0.69
EP_NOHSDP	0.09	0.55	0.75	0.62
EP_AGE65	0.02	-0.12	-0.02	-0.13
EP_AGE17	-0.10	-0.01	0.04	-0.04
EP_DISABL	0.21	0.34	0.65	0.41
EP_SNGPNT	0.25	0.37	0.60	0.42
EP_MINRTY	0.76	0.30	0.64	0.40
EP_LIMENG	0.78	0.07	0.14	0.08
EP_MUNIT	0.43	0.41	0.24	0.22
EP_MOBILE	-0.47	0.04	-0.07	-0.01
EP_CROWD	0.15	0.61	0.55	0.34
EP_NOVEH	0.31	0.60	0.74	0.58
EP_GROUPQ	0.16	0.64	0.44	0.32
RPL_THEME1	0.28	0.56	0.87	0.66
RPL_THEME2	0.20	0.31	0.67	0.39
RPL_THEME3	***	0.24	0.50	0.30
RPL_THEME4	0.24	***	0.78	0.58
RPL_THEMES	0.50	0.78	***	0.72
EP_UNINSUR	0.30	0.58	0.72	***

(Table Cont'd)

Appendix J. Python code for Neural Network ED Model

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.neural_network import MLPClassifier
from sklearn import metrics
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
dataset = pd.read_csv('//entfilesrvr/som-br$/New
Transfer/Population Health/RamyaKrishna
Tummala/EDVISITS_July23/ED_Visits_July30/EDVisits_SPSS/ED_HighUt
ilizers_0801/LA_2015_2019_EDHighUtilizers_Sampling_Python_Train_
Test.csv')
dataset.dtypes
print("Datset Shape:")
print(dataset.shape)
X = dataset.iloc[:, 0:19].values
y = dataset.iloc[:, 19].values
from sklearn.model_selection import KFold
accuracy=[]
F1_Score=[]
ROC_AUC=[]
ROC_AUC_PVP=[]
kf = KFold(n_splits=10, random_state=12, shuffle=True)
print("10-fold Cross Validation")
for train_index, test_index in kf.split(X,y):
    print("TRAIN:", train_index, "TEST:", test_index)
    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y[train_index], y[test_index]
    scaler.fit(X_train)
    X_train = scaler.transform(X_train)
    X_test = scaler.transform(X_test)
    Clf = MLPClassifier(hidden_layer_sizes=(20,20,20),
  activation= 'relu', solver='sgd', alpha= 0.3,learning_rate =
  'adaptive', learning_rate_init = 0.3, momentum=0.4,
  max_iter=1000)
    Clf.fit(X_train, y_train)
    y_predk = Clf.predict(X_test)
    df = pd.DataFrame({'Actual': y_test, 'Predicted': y_predk})
    print(df.head(25))
    print("***Evaluation Metrics for testing dataset***")
    Confusion_Matrix = metrics.confusion_matrix(y_test, y_predk)
    print("Confusion Matrix:")
    print(metrics.confusion_matrix(y_test, y_predk))
    print("Accuracy Score:",metrics.accuracy_score(y_test,
y_predk))
```

```
    print("Misclassification Rate (Classification error):", 1 -
metrics.accuracy_score(y_test, y_predk))
    print("Sensitivity:", metrics.recall_score(y_test, y_predk))
    print("Specificity:", (Confusion_Matrix[0][0] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
    print("False Positive Rate:", (Confusion_Matrix[0][1] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
    print("Precision:", metrics.precision_score(y_test,
y_predk))
    print("F1 Score:" ,metrics.f1_score(y_test, y_predk))
    print("ROC_AUC_SCORE of Predicted
Value:",metrics.roc_auc_score(y_test, y_predk))
    y_pred_probk = Clf.predict_proba(X_test)[:, 1]
    print("ROC_AUC_SCORE OF Predictied Value Probability:",
metrics.roc_auc_score(y_test, y_pred_probk))
    Accuracy_Score = metrics.accuracy_score(y_test, y_predk)
    F1_score = metrics.f1_score(y_test, y_predk)
    ROC_AUC_SCORE=metrics.roc_auc_score(y_test, y_predk)
    ROC_AUC_SCORE_PVP=metrics.roc_auc_score(y_test,
y_pred_probk)
    accuracy.append(Accuracy_Score)
    F1_Score.append(F1_score)
    ROC_AUC.append(ROC_AUC_SCORE)
    ROC_AUC_PVP.append(ROC_AUC_SCORE_PVP)
Accuracy=np.array(accuracy).mean()
F1_SCORE=np.array(F1_Score).mean()
roc_auc=np.array(ROC_AUC).mean()
roc_auc_pvp=np.array(ROC_AUC_PVP).mean()
print("Average Accuracy for 10 Folds:", Accuracy)
print("Average F1-Score for 10 Folds:", F1_SCORE)
print("Average ROC_AUC for 10 Folds:", roc_auc)
print("Average ROC_AUC_PVP for 10 Folds:", roc_auc_pvp)
dataset_Validation = pd.read_csv('//entfilesrvr/som-br$/New
Transfer/Population Health/RamyaKrishna
Tummala/EDVISITS_July23/ED_Visits_July30/EDVisits_SPSS/ED_HighUt
ilizers_0801/LA_2015_2019_EDHighUtilizers_Sampling_Python_Valida
tion.csv')
dataset_Validation.dtypes
print("Datset Shape:")
print(dataset_Validation.shape)
X_valid = dataset_Validation.iloc[:, 0:19].values
y_valid = dataset_Validation.iloc[:, 19].values
X_valid = scaler.transform(X_valid)
y_pred = Clf.predict(X_valid)
df = pd.DataFrame({'Actual': y_valid, 'Predicted': y_pred})
print(df.head(25))
print("***Evaluation Metrics of Validation dataset***")
```

```
from sklearn import metrics
Confusion_Matrix = metrics.confusion_matrix(y_valid, y_pred)
print("Confusion Matrix:")
print(metrics.confusion_matrix(y_valid, y_pred))
print("Accuracy Score:",metrics.accuracy_score(y_valid, y_pred))
print("Misclassification Rate (Classification error):", 1 -
metrics.accuracy_score(y_valid, y_pred))
print("Sensitivity:", metrics.recall_score(y_valid, y_pred))
print("Specificity:", (Confusion_Matrix[0][0] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
print("False Positive Rate:", (Confusion_Matrix[0][1] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
print("Precision:", metrics.precision_score(y_valid, y_pred))
print("F1 Score:" ,metrics.f1_score(y_valid, y_pred))
fpr, tpr, thresholds = metrics.roc_curve(y_valid, y_pred)
import matplotlib.pyplot as plt
plt.plot(fpr, tpr)
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.0])
plt.rcParams['font.size'] = 12
plt.title('ROC curve for ED High Utilizers classifier')
plt.xlabel('False Positive Rate (1 - Specificity)')
plt.ylabel('True Positive Rate (Sensitivity)')
plt.grid(True)
plt.show()
print("ROC_AUC_SCORE of Predicted
Value:",metrics.roc_auc_score(y_valid, y_pred))
y_pred_prob = Clf.predict_proba(X_valid)[:, 1]
fpr, tpr, thresholds = metrics.roc_curve(y_valid, y_pred_prob)
plt.plot(fpr, tpr)
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.0])
plt.rcParams['font.size'] = 12
plt.title('ROC curve for ED High Utilizers classifier')
plt.xlabel('False Positive Rate (1 - Specificity)')
plt.ylabel('True Positive Rate (Sensitivity)')
plt.grid(True)
plt.show()
print("ROC_AUC_SCORE OF Predictied Value Probability:",
metrics.roc_auc_score(y_valid, y_pred_prob))
```

Appendix K. Python code for Random Forest ED Model

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.ensemble import RandomForestClassifier
from sklearn import metrics
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
dataset = pd.read_csv('//entfilesrvr/som-br$/New
Transfer/Population Health/RamyaKrishna
Tummala/EDVISITS_July23/ED_Visits_July30/EDVisits_SPSS/ED_HighUt
ilizers_0801/LA_2015_2019_EDHighUtilizers_Sampling_Python_Train_
Test.csv')
dataset.dtypes
print("Datset Shape:")
print(dataset.shape)
X = dataset.iloc[:, 0:19].values
y = dataset.iloc[:, 19].values
from sklearn.model_selection import KFold
accuracy=[]
F1_Score=[]
ROC_AUC=[]
ROC_AUC_PVP=[]
kf = KFold(n_splits=10, random_state=12, shuffle=True)
print("10-fold Cross Validation")
for train_index, test_index in kf.split(X,y):
    print("TRAIN:", train_index, "TEST:", test_index)
    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y[train_index], y[test_index]
    scaler.fit(X_train)
    X_train = scaler.transform(X_train)
    X_test = scaler.transform(X_test)
    classifier =
RandomForestClassifier(n_estimators=1800,criterion='entropy',max
_depth=90,min_samples_split=2,min_samples_leaf=1,max_features='a
uto',bootstrap='True', random_state=12)
    classifier.fit(X_train, y_train)
    y_predk = classifier.predict(X_test)
    df = pd.DataFrame({'Actual': y_test, 'Predicted': y_predk})
    print(df.head(25))
    print("***Evaluation Metrics for testing dataset***")
    Confusion_Matrix = metrics.confusion_matrix(y_test, y_predk)
    print("Confusion Matrix:")
    print(metrics.confusion_matrix(y_test, y_predk))
    print("Accuracy Score:",metrics.accuracy_score(y_test,
y_predk))
```

```
    print("Misclassification Rate (Classification error):", 1 -
metrics.accuracy_score(y_test, y_predk))
    print("Sensitivity:", metrics.recall_score(y_test, y_predk))
    print("Specificity:", (Confusion_Matrix[0][0] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
    print("False Positive Rate:", (Confusion_Matrix[0][1] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
    print("Precision:", metrics.precision_score(y_test,
y_predk))
    print("F1 Score:" ,metrics.f1_score(y_test, y_predk))
    print("ROC_AUC_SCORE of Predicted
Value:",metrics.roc_auc_score(y_test, y_predk))
    y_pred_probk = classifier.predict_proba(X_test)[:, 1]
    print("ROC_AUC_SCORE OF Predictied Value Probability:",
metrics.roc_auc_score(y_test, y_pred_probk))
    Accuracy_Score = metrics.accuracy_score(y_test, y_predk)
    F1_score = metrics.f1_score(y_test, y_predk)
    ROC_AUC_SCORE=metrics.roc_auc_score(y_test, y_predk)
    ROC_AUC_SCORE_PVP=metrics.roc_auc_score(y_test,
y_pred_probk)
    accuracy.append(Accuracy_Score)
    F1_Score.append(F1_score)
    ROC_AUC.append(ROC_AUC_SCORE)
    ROC_AUC_PVP.append(ROC_AUC_SCORE_PVP)
Accuracy=np.array(accuracy).mean()
F1_SCORE=np.array(F1_Score).mean()
roc_auc=np.array(ROC_AUC).mean()
roc_auc_pvp=np.array(ROC_AUC_PVP).mean()
print("Average Accuracy for 10 Folds:", Accuracy)
print("Average F1-Score for 10 Folds:", F1_SCORE)
print("Average ROC_AUC for 10 Folds:", roc_auc)
print("Average ROC_AUC_PVP for 10 Folds:", roc_auc_pvp)
dataset_Validation = pd.read_csv('//entfilesrvr/som-br$/New
Transfer/Population Health/RamyaKrishna
Tummala/EDVISITS_July23/ED_Visits_July30/EDVisits_SPSS/ED_HighUt
ilizers_0801/LA_2015_2019_EDHighUtilizers_Sampling_Python_Valida
tion.csv')
dataset_Validation.dtypes
print("Datset Shape:")
print(dataset_Validation.shape)
X_valid = dataset_Validation.iloc[:, 0:19].values
y_valid = dataset_Validation.iloc[:, 19].values
X_valid = scaler.transform(X_valid)
y_pred = classifier.predict(X_valid)
df = pd.DataFrame({'Actual': y_valid, 'Predicted': y_pred})
print(df.head(25))
print("***Evaluation Metrics of Validation dataset***")
```

```
from sklearn import metrics
Confusion_Matrix = metrics.confusion_matrix(y_valid, y_pred)
print("Confusion Matrix:")
print(metrics.confusion_matrix(y_valid, y_pred))
print("Accuracy Score:",metrics.accuracy_score(y_valid, y_pred))
print("Misclassification Rate (Classification error):", 1 -
metrics.accuracy_score(y_valid, y_pred))
print("Sensitivity:", metrics.recall_score(y_valid, y_pred))
print("Specificity:", (Confusion_Matrix[0][0] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
print("False Positive Rate:", (Confusion_Matrix[0][1] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
print("Precision:", metrics.precision_score(y_valid, y_pred))
print("F1 Score:" ,metrics.f1_score(y_valid, y_pred))
fpr, tpr, thresholds = metrics.roc_curve(y_valid, y_pred)
import matplotlib.pyplot as plt
plt.plot(fpr, tpr)
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.0])
plt.rcParams['font.size'] = 12
plt.title('ROC curve for ED High Utilizers classifier')
plt.xlabel('False Positive Rate (1 - Specificity)')
plt.ylabel('True Positive Rate (Sensitivity)')
plt.grid(True)
plt.show()
print("ROC_AUC_SCORE of Predicted
Value:",metrics.roc_auc_score(y_valid, y_pred))
y_pred_prob = classifier.predict_proba(X_valid)[:, 1]
fpr, tpr, thresholds = metrics.roc_curve(y_valid, y_pred_prob)
plt.plot(fpr, tpr)
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.0])
plt.rcParams['font.size'] = 12
plt.title('ROC curve for ED High Utilizers classifier')
plt.xlabel('False Positive Rate (1 - Specificity)')
plt.ylabel('True Positive Rate (Sensitivity)')
plt.grid(True)
plt.show()
print("ROC_AUC_SCORE OF Predictied Value Probability:",
metrics.roc_auc_score(y_valid, y_pred_prob))
```

Appendix L. Python code for XGBoost ED Model

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from xgboost import XGBClassifier
from sklearn import metrics
from sklearn.preprocessing import StandardScaler
scaler = StandardScaler()
dataset = pd.read_csv('//entfilesrvr/som-br$/New
Transfer/Population Health/RamyaKrishna
Tummala/EDVISITS_July23/ED_Visits_July30/EDVisits_SPSS/ED_HighUt
ilizers_0801/LA_2015_2019_EDHighUtilizers_Sampling_Python_Train_
Test.csv')
dataset.dtypes
print("Datset Shape:")
print(dataset.shape)
X = dataset.iloc[:, 0:19].values
y = dataset.iloc[:, 19].values
from sklearn.model_selection import KFold
accuracy=[]
F1_Score=[]
ROC_AUC=[]
ROC_AUC_PVP=[]
kf = KFold(n_splits=10, random_state=12, shuffle=True)
print("10-fold Cross Validation")
for train_index, test_index in kf.split(X,y):
    print("TRAIN:", train_index, "TEST:", test_index)
    X_train, X_test = X[train_index], X[test_index]
    y_train, y_test = y[train_index], y[test_index]
    scaler.fit(X_train)
    X_train = scaler.transform(X_train)
    X_test = scaler.transform(X_test)
    xgb_clf = XGBClassifier(base_score=0.5, colsample_bylevel=1,
                colsample_bynode=1, colsample_bytree=0.9,
gamma=0.1, gpu_id=-1,
                importance_type='gain',
                learning_rate=0.05, max_delta_step=0,
max_depth=12,
                min_child_weight=5,
                n_estimators=100, num_parallel_tree=1,
                objective='binary:logistic', scale_pos_weight=1,
subsample=1)
    xgb_clf.fit(X_train, y_train)
    y_predk = xgb_clf.predict(X_test)
    df = pd.DataFrame({'Actual': y_test, 'Predicted': y_predk})
    print(df.head(25))
```

```
    print("***Evaluation Metrics for testing dataset***")
    Confusion_Matrix = metrics.confusion_matrix(y_test, y_predk)
    print("Confusion Matrix:")
    print(metrics.confusion_matrix(y_test, y_predk))
    print("Accuracy Score:",metrics.accuracy_score(y_test,
y_predk))
    print("Misclassification Rate (Classification error):", 1 -
metrics.accuracy_score(y_test, y_predk))
    print("Sensitivity:", metrics.recall_score(y_test, y_predk))
    print("Specificity:", (Confusion_Matrix[0][0] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
    print("False Positive Rate:", (Confusion_Matrix[0][1] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
    print("Precision:", metrics.precision_score(y_test,
y_predk))
    print("F1 Score:" ,metrics.f1_score(y_test, y_predk))
    print("ROC_AUC_SCORE of Predicted
Value:",metrics.roc_auc_score(y_test, y_predk))
    y_pred_probk = xgb_clf.predict_proba(X_test)[:, 1]
    print("ROC_AUC_SCORE OF Predictied Value Probability:",
metrics.roc_auc_score(y_test, y_pred_probk))
    Accuracy_Score = metrics.accuracy_score(y_test, y_predk)
    F1_score = metrics.f1_score(y_test, y_predk)
    ROC_AUC_SCORE=metrics.roc_auc_score(y_test, y_predk)
    ROC_AUC_SCORE_PVP=metrics.roc_auc_score(y_test,
y_pred_probk)
    accuracy.append(Accuracy_Score)
    F1_Score.append(F1_score)
    ROC_AUC.append(ROC_AUC_SCORE)
    ROC_AUC_PVP.append(ROC_AUC_SCORE_PVP)
Accuracy=np.array(accuracy).mean()
F1_SCORE=np.array(F1_Score).mean()
roc_auc=np.array(ROC_AUC).mean()
roc_auc_pvp=np.array(ROC_AUC_PVP).mean()
print("Average Accuracy for 10 Folds:", Accuracy)
print("Average F1-Score for 10 Folds:", F1_SCORE)
print("Average ROC_AUC for 10 Folds:", roc_auc)
print("Average ROC_AUC_PVP for 10 Folds:", roc_auc_pvp)
dataset_Validation = pd.read_csv('//entfilesrvr/som-br$/New
Transfer/Population Health/RamyaKrishna
Tummala/EDVISITS_July23/ED_Visits_July30/EDVisits_SPSS/ED_HighUt
ilizers_0801/LA_2015_2019_EDHighUtilizers_Sampling_Python_Valida
tion.csv')
dataset_Validation.dtypes
print("Datset Shape:")
print(dataset_Validation.shape)
X_valid = dataset_Validation.iloc[:, 0:19].values
```

```
y_valid = dataset_Validation.iloc[:, 19].values
X_valid = scaler.transform(X_valid)
y_pred = xgb_clf.predict(X_valid)
df = pd.DataFrame({'Actual': y_valid, 'Predicted': y_pred})
print(df.head(25))
print("***Evaluation Metrics of Validation dataset***")
from sklearn import metrics
Confusion_Matrix = metrics.confusion_matrix(y_valid, y_pred)
print("Confusion Matrix:")
print(metrics.confusion_matrix(y_valid, y_pred))
print("Accuracy Score:",metrics.accuracy_score(y_valid, y_pred))
print("Misclassification Rate (Classification error):", 1 -
metrics.accuracy_score(y_valid, y_pred))
print("Sensitivity:", metrics.recall_score(y_valid, y_pred))
print("Specificity:", (Confusion_Matrix[0][0] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
print("False Positive Rate:", (Confusion_Matrix[0][1] /
(Confusion_Matrix[0][0] + Confusion_Matrix[0][1])))
print("Precision:", metrics.precision_score(y_valid, y_pred))
print("F1 Score:" ,metrics.f1_score(y_valid, y_pred))
fpr, tpr, thresholds = metrics.roc_curve(y_valid, y_pred)
import matplotlib.pyplot as plt
plt.plot(fpr, tpr)
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.0])
plt.rcParams['font.size'] = 12
plt.title('ROC curve for ED High Utilizers classifier')
plt.xlabel('False Positive Rate (1 - Specificity)')
plt.ylabel('True Positive Rate (Sensitivity)')
plt.grid(True)
plt.show()
print("ROC_AUC_SCORE of Predicted
Value:",metrics.roc_auc_score(y_valid, y_pred))
y_pred_prob = xgb_clf.predict_proba(X_valid)[:, 1]
fpr, tpr, thresholds = metrics.roc_curve(y_valid, y_pred_prob)
plt.plot(fpr, tpr)
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.0])
plt.rcParams['font.size'] = 12
plt.title('ROC curve for ED High Utilizers classifier')
plt.xlabel('False Positive Rate (1 - Specificity)')
plt.ylabel('True Positive Rate (Sensitivity)')
plt.grid(True)
plt.show()
print("ROC_AUC_SCORE   OF   Predictied   Value   Probability:",
metrics.roc_auc_score(y_valid, y_pred_prob))
```

Appendix M. Hot Spot Analysis FMOLHS ED High Utilizers Count By Census Tracts across LA

Census Tract ID	Parish	ED High Utilizers Count	Confidence Result
1	East Baton Rouge	58	Hot Spot 99%
10	East Baton Rouge	88	Hot Spot 99%
11.02	East Baton Rouge	67	Hot Spot 99%
11.03	East Baton Rouge	77	Hot Spot 99%
11.04	East Baton Rouge	174	Hot Spot 99%
16	East Baton Rouge	59	Hot Spot 99%
17	East Baton Rouge	54	Hot Spot 99%
18	East Baton Rouge	75	Hot Spot 99%
19	East Baton Rouge	11	Hot Spot 99%
2	East Baton Rouge	207	Hot Spot 99%
20	East Baton Rouge	12	Hot Spot 99%
201	West Baton Rouge	58	Hot Spot 99%
202	West Baton Rouge	48	Hot Spot 99%
203	West Baton Rouge	103	Hot Spot 99%
204.01	West Baton Rouge	36	Hot Spot 99%
204.02	West Baton Rouge	33	Hot Spot 99%
22	East Baton Rouge	37	Hot Spot 99%
23	East Baton Rouge	8	Hot Spot 99%
24	East Baton Rouge	71	Hot Spot 99%
25	East Baton Rouge	66	Hot Spot 99%
26.01	East Baton Rouge	14	Hot Spot 99%
26.02	East Baton Rouge	78	Hot Spot 99%
27	East Baton Rouge	49	Hot Spot 99%
28.01	East Baton Rouge	5	Hot Spot 99%
28.02	East Baton Rouge	4	Hot Spot 99%
3	East Baton Rouge	123	Hot Spot 99%
30	East Baton Rouge	95	Hot Spot 99%
301.01	Ascension	49	Hot Spot 99%
301.02	Ascension	25	Hot Spot 99%
301.03	Ascension	32	Hot Spot 99%
302.03	Ascension	47	Hot Spot 99%
302.04	Ascension	79	Hot Spot 99%
302.05	Ascension	59	Hot Spot 99%
302.06	Ascension	83	Hot Spot 99%
303	Ascension	138	Hot Spot 99%
304.01	Ascension	87	Hot Spot 99%
304.02	Ascension	125	Hot Spot 99%
305	Ascension	66	Hot Spot 99%

(Table Cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Confidence Result
306	Ascension	80	Hot Spot 99%
309	Ascension	18	Hot Spot 99%
31.01	East Baton Rouge	83	Hot Spot 99%
31.03	East Baton Rouge	31	Hot Spot 99%
310	Ascension	28	Hot Spot 99%
32.01	East Baton Rouge	53	Hot Spot 99%
32.02	East Baton Rouge	26	Hot Spot 99%
33	East Baton Rouge	86	Hot Spot 99%
34	East Baton Rouge	197	Hot Spot 99%
35.01	East Baton Rouge	31	Hot Spot 99%
35.04	East Baton Rouge	140	Hot Spot 99%
35.05	East Baton Rouge	117	Hot Spot 99%
35.06	East Baton Rouge	82	Hot Spot 99%
35.07	East Baton Rouge	65	Hot Spot 99%
36.01	East Baton Rouge	45	Hot Spot 99%
36.03	East Baton Rouge	43	Hot Spot 99%
36.04	East Baton Rouge	137	Hot Spot 99%
37.01	East Baton Rouge	46	Hot Spot 99%
37.02	East Baton Rouge	75	Hot Spot 99%
37.03	East Baton Rouge	61	Hot Spot 99%
38.01	East Baton Rouge	128	Hot Spot 99%
38.02	East Baton Rouge	43	Hot Spot 99%
38.04	East Baton Rouge	23	Hot Spot 99%
38.05	East Baton Rouge	86	Hot Spot 99%
39.04	East Baton Rouge	63	Hot Spot 99%
39.06	East Baton Rouge	43	Hot Spot 99%
39.07	East Baton Rouge	36	Hot Spot 99%
39.08	East Baton Rouge	31	Hot Spot 99%
39.09	East Baton Rouge	51	Hot Spot 99%
39.1	East Baton Rouge	49	Hot Spot 99%
4	East Baton Rouge	90	Hot Spot 99%
40.05	East Baton Rouge	86	Hot Spot 99%
40.06	East Baton Rouge	34	Hot Spot 99%
40.09	East Baton Rouge	57	Hot Spot 99%
40.1	East Baton Rouge	24	Hot Spot 99%
40.11	East Baton Rouge	141	Hot Spot 99%
40.13	East Baton Rouge	79	Hot Spot 99%
40.14	East Baton Rouge	55	Hot Spot 99%
40.15	East Baton Rouge	167	Hot Spot 99%
40.16	East Baton Rouge	49	Hot Spot 99%
401	Livingston	74	Hot Spot 99%

(Table Cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Confidence Result
402.01	Livingston	411	Hot Spot 99%
402.02	Livingston	143	Hot Spot 99%
403.01	Livingston	192	Hot Spot 99%
403.03	Livingston	124	Hot Spot 99%
403.04	Livingston	155	Hot Spot 99%
404.01	Livingston	256	Hot Spot 99%
404.02	Livingston	280	Hot Spot 99%
405	Livingston	250	Hot Spot 99%
406	Livingston	160	Hot Spot 99%
407	Livingston	71	Hot Spot 99%
408.02	Livingston	114	Hot Spot 99%
408.04	Livingston	673	Hot Spot 99%
408.05	Livingston	103	Hot Spot 99%
408.06	Livingston	162	Hot Spot 99%
409.01	Livingston	48	Hot Spot 99%
409.02	Livingston	38	Hot Spot 99%
42.01	East Baton Rouge	61	Hot Spot 99%
42.03	East Baton Rouge	39	Hot Spot 99%
42.04	East Baton Rouge	51	Hot Spot 99%
42.05	East Baton Rouge	64	Hot Spot 99%
43.01	East Baton Rouge	31	Hot Spot 99%
43.02	East Baton Rouge	19	Hot Spot 99%
44.01	East Baton Rouge	22	Hot Spot 99%
44.02	East Baton Rouge	29	Hot Spot 99%
44.03	East Baton Rouge	21	Hot Spot 99%
45.03	East Baton Rouge	76	Hot Spot 99%
45.04	East Baton Rouge	36	Hot Spot 99%
45.05	East Baton Rouge	69	Hot Spot 99%
45.07	East Baton Rouge	17	Hot Spot 99%
45.08	East Baton Rouge	16	Hot Spot 99%
45.09	East Baton Rouge	53	Hot Spot 99%
45.1	East Baton Rouge	41	Hot Spot 99%
46.02	East Baton Rouge	33	Hot Spot 99%
46.03	East Baton Rouge	15	Hot Spot 99%
46.04	East Baton Rouge	25	Hot Spot 99%
47	East Baton Rouge	27	Hot Spot 99%
48	East Baton Rouge	20	Hot Spot 99%
49	East Baton Rouge	18	Hot Spot 99%
5	East Baton Rouge	132	Hot Spot 99%
50	East Baton Rouge	4	Hot Spot 99%
51	East Baton Rouge	76	Hot Spot 99%

(Table Cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Confidence Result
52	East Baton Rouge	65	Hot Spot 99%
53	East Baton Rouge	202	Hot Spot 99%
6.01	East Baton Rouge	71	Hot Spot 99%
6.02	East Baton Rouge	116	Hot Spot 99%
7.01	East Baton Rouge	63	Hot Spot 99%
7.02	East Baton Rouge	74	Hot Spot 99%
9	East Baton Rouge	108	Hot Spot 99%
9512	St. Helena	31	Hot Spot 99%
9514	East Feliciana	9	Hot Spot 99%
9515.01	East Feliciana	4	Hot Spot 99%
9515.02	East Feliciana	28	Hot Spot 99%
9516	East Feliciana	14	Hot Spot 99%
9518	West Feliciana	13	Hot Spot 99%
9519	Pointe Coupee	24	Hot Spot 99%
9521	Pointe Coupee	23	Hot Spot 99%
9522	Pointe Coupee	16	Hot Spot 99%
9523	Pointe Coupee	7	Hot Spot 99%
9526	Iberville	36	Hot Spot 99%
9527	Iberville	35	Hot Spot 99%
9529	Iberville	26	Hot Spot 99%
9530	Iberville	1	Hot Spot 99%
9531.01	Iberville	20	Hot Spot 99%
9531.02	Iberville	15	Hot Spot 99%
9532	Iberville	78	Hot Spot 99%
9536	Tangipahoa	1	Hot Spot 99%
9538	Tangipahoa	4	Hot Spot 99%
9539	Tangipahoa	8	Hot Spot 99%
9540.01	Tangipahoa	2	Hot Spot 99%
9540.02	Tangipahoa	5	Hot Spot 99%
9541.01	Tangipahoa	8	Hot Spot 99%
9541.02	Tangipahoa	3	Hot Spot 99%
9543	Tangipahoa	6	Hot Spot 99%
9544	Tangipahoa	1	Hot Spot 99%
9545.01	Tangipahoa	2	Hot Spot 99%
9545.02	Tangipahoa	1	Hot Spot 99%
404	St. James	16	Hot Spot 95%
9524	Pointe Coupee	14	Hot Spot 95%
9547	Tangipahoa	1	Hot Spot 90%
216.02	Lafourche	1	Cold Spot 90%
401.02	St. Tammany	2	Cold Spot 90%

(Table Cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Confidence Result
407.05	St. Tammany	1	Cold Spot 90%
708	St. John the Baptist	1	Cold Spot 90%
9506	Washington	1	Cold Spot 90%
9535	Tangipahoa	3	Cold Spot 90%
9602	Acadia	1	Cold Spot 90%
9604	St. Landry	2	Cold Spot 90%
9610	St. Landry	1	Cold Spot 90%
1.01	Terrebonne	1	Cold Spot 95%
1.02	Terrebonne	2	Cold Spot 95%
12	Lafayette	2	Cold Spot 95%
14.02	Lafayette	1	Cold Spot 95%
14.05	Lafayette	1	Cold Spot 95%
14.06	Lafayette	2	Cold Spot 95%
14.09	Lafayette	1	Cold Spot 95%
14.11	Lafayette	1	Cold Spot 95%
17	Terrebonne	1	Cold Spot 95%
2.01	Terrebonne	1	Cold Spot 95%
2.02	Terrebonne	2	Cold Spot 95%
202	St. Martin	2	Cold Spot 95%
204	St. Martin	1	Cold Spot 95%
205	Lafourche	1	Cold Spot 95%
205.02	St. Martin	1	Cold Spot 95%
207.03	Lafourche	1	Cold Spot 95%
217	Lafourche	1	Cold Spot 95%
303.01	Iberia	1	Cold Spot 95%
311	Iberia	1	Cold Spot 95%
402	St. Mary	1	Cold Spot 95%
403.05	St. Tammany	1	Cold Spot 95%
404	St. Tammany	1	Cold Spot 95%
405.02	St. Tammany	1	Cold Spot 95%
411.03	St. Tammany	1	Cold Spot 95%
5	Terrebonne	2	Cold Spot 95%
711	St. John the Baptist	1	Cold Spot 95%
9507	Vermilion	1	Cold Spot 95%
9605	Acadia	2	Cold Spot 95%
9607	Acadia	1	Cold Spot 95%
9610	Acadia	1	Cold Spot 95%
9613	St. Landry	1	Cold Spot 95%
9616	St. Landry	2	Cold Spot 95%
9618	St. Landry	1	Cold Spot 95%
1	Lafayette	5	Cold Spot 99%
10.01	Lafayette	2	Cold Spot 99%

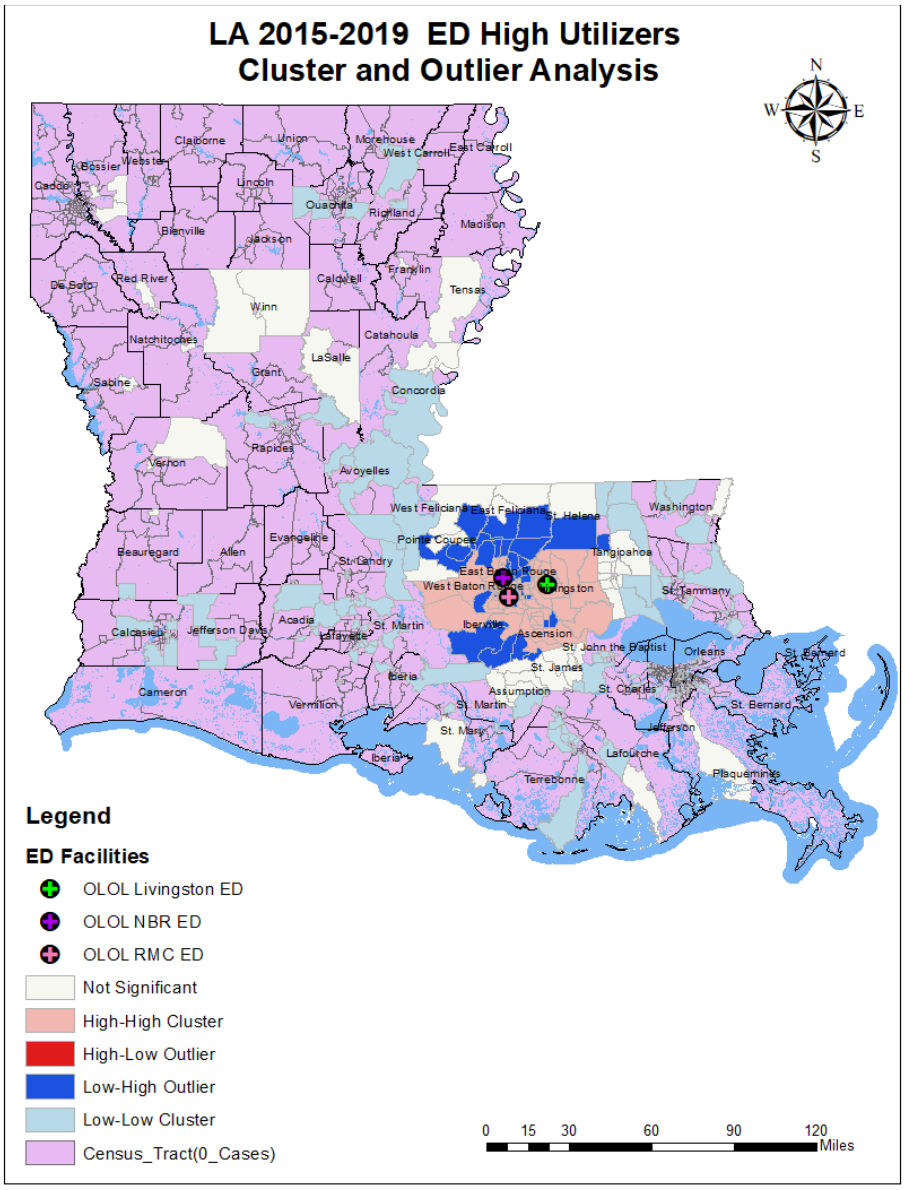
(Table Cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Confidence Result
10.01	Lafayette	2	Cold Spot 99%
123	Orleans	1	Cold Spot 99%
13.02	Orleans	1	Cold Spot 99%
135	Orleans	1	Cold Spot 99%
138	Orleans	1	Cold Spot 99%
14.01	Orleans	1	Cold Spot 99%
141	Orleans	1	Cold Spot 99%
17.22	Orleans	1	Cold Spot 99%
17.23	Orleans	1	Cold Spot 99%
17.25	Orleans	1	Cold Spot 99%
17.4	Orleans	2	Cold Spot 99%
17.45	Orleans	1	Cold Spot 99%
17.46	Orleans	1	Cold Spot 99%
202.02	Jefferson	1	Cold Spot 99%
205.08	Jefferson	1	Cold Spot 99%
205.14	Jefferson	1	Cold Spot 99%
206	Jefferson	3	Cold Spot 99%
21.02	Lafayette	1	Cold Spot 99%
22	Lafayette	1	Cold Spot 99%
221.02	Jefferson	1	Cold Spot 99%
231	Jefferson	1	Cold Spot 99%
24.01	Orleans	1	Cold Spot 99%
246	Jefferson	1	Cold Spot 99%
251.02	Jefferson	2	Cold Spot 99%
252.02	Jefferson	1	Cold Spot 99%
261	Jefferson	1	Cold Spot 99%
266	Jefferson	1	Cold Spot 99%
27	Orleans	1	Cold Spot 99%
270	Jefferson	1	Cold Spot 99%
276.01	Jefferson	2	Cold Spot 99%
277.01	Jefferson	2	Cold Spot 99%
278.07	Jefferson	1	Cold Spot 99%
30	Orleans	1	Cold Spot 99%
33.03	Orleans	1	Cold Spot 99%
37.01	Orleans	1	Cold Spot 99%
37.02	Orleans	1	Cold Spot 99%
408.01	St. Tammany	2	Cold Spot 99%
408.03	St. Tammany	1	Cold Spot 99%
412.12	St. Tammany	1	Cold Spot 99%
45	Orleans	1	Cold Spot 99%
6.07	Orleans	1	Cold Spot 99%

(Table Cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Confidence Result
6.11	Orleans	1	Cold Spot 99%
6.12	Orleans	1	Cold Spot 99%
6.13	Orleans	1	Cold Spot 99%
60	Orleans	2	Cold Spot 99%
623.01	St. Charles	1	Cold Spot 99%
629	St. Charles	1	Cold Spot 99%
63	Orleans	1	Cold Spot 99%
702	St. John the Baptist	2	Cold Spot 99%
703	St. John the Baptist	1	Cold Spot 99%
85	Orleans	1	Cold Spot 99%
9548	Tangipahoa	3	Cold Spot 99%

Appendix N. Cluster and Outlier Analysis of FMOLHS LA ED High Utilizers



ED High Utilizers

Appendix O. FMOLHS LA ED High Utilizers Map Summary Tables of Cluster and Outlier Analysis

Results	Area (Kilometer Square)	# of Census Tracts	# of ED High Utilizers	Total Population	# ED High Utilizers/total pop x 1000
Not Significant	12518.64	50	241	239966	1.00
Cluster High-High	4263.49	99	9723	591627	16.43
Outlier Low-High	3655.92	43	845	212677	3.97
Cluster Low-Low	12713.66	135	186	676216	0.28

Appendix P. Cluster and Outlier Analysis Percentage Averages of SVI and ACS Data Based on FMOLHS LA ED High Utilizers

LEGEND
High Average Values

Results /SVI data % averages	High-High	Low-High	Low-Low
Poverty	20.53	15.28	22.78
Unemployment	7.80	6.31	8.00
Per Capita Income	27071.31	32639.81	24797.44
No High School Dip	14.13	10.60	18.42
Age 65	12.71	15.47	14.43
Age 17	23.97	21.24	23.64
Disability	15.80	14.45	15.52
Single Parent	12.03	8.96	12.35
Minority	54.61	36.77	49.01
Limited English Proficiency	1.61	0.97	1.65
Housing 10 or More	12.71	8.55	7.14
Mobile Homes	9.68	12.12	12.67
Crowded Housing	2.92	2.48	2.45
No Vehicle	8.86	5.95	11.63
Institutionalized Group Quarters	2.01	4.85	2.86
Socioeconomic Rank	0.45	0.33	0.56
Household Composition Rank	0.46	0.42	0.54
Minority/Language Rank	0.55	0.41	0.57
Housing/Transportation Rank	0.55	0.41	0.51
Overall Rank	0.49	0.35	0.56
Uninsured	11.40	7.86	11.62

Appendix Q. Cluster and Outlier Analysis FMOLHS ED High Utilizer Counts By Census Tracts across LA

Census Tract ID	Parish	ED High Utilizers Count	Result
1	East Baton Rouge	58	HH
1	Lafayette	5	LL
1.01	Terrebonne	1	LL
1.02	Terrebonne	2	LL
10	East Baton Rouge	88	HH
10.01	Lafayette	2	LL
101	Rapides	1	LL
104	Ouachita	1	LL
105	Rapides	1	LL
105.04	Ouachita	1	LL
106.03	Ouachita	1	LL
108	Ouachita	1	LL
11.02	East Baton Rouge	67	HH
11.03	East Baton Rouge	77	HH
11.04	East Baton Rouge	174	HH
12	Lafayette	2	LL
12.02	Calcasieu	1	LL
120	Rapides	1	LL
122	Rapides	1	LL
123	Orleans	1	LL
13	Terrebonne	1	LL
13.02	Orleans	1	LL
135	Orleans	1	LL
138	Orleans	1	LL
14	Calcasieu	1	LL
14.01	Orleans	1	LL
14.02	Lafayette	1	LL
14.05	Lafayette	1	LL
14.06	Lafayette	2	LL
14.09	Lafayette	1	LL
14.11	Lafayette	1	LL
141	Orleans	1	LL
16	East Baton Rouge	59	HH
17	Calcasieu	2	LL
17	East Baton Rouge	54	HH
17	Terrebonne	1	LL
17.22	Orleans	1	LL
17.23	Orleans	1	LL

(Table cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Result
17.25	Orleans	1	LL
17.4	Orleans	2	LL
17.45	Orleans	1	LL
17.46	Orleans	1	LL
18	East Baton Rouge	75	HH
18.01	Calcasieu	1	LL
19	East Baton Rouge	11	LH
2	East Baton Rouge	207	HH
2	Jefferson Davis	1	LL
2.01	Terrebonne	1	LL
2.02	Terrebonne	2	LL
20	East Baton Rouge	12	LH
201	West Baton Rouge	58	HH
202	St. Martin	2	LL
202	West Baton Rouge	48	HH
202.02	Jefferson	1	LL
203	West Baton Rouge	103	HH
204	St. Martin	1	LL
204.01	West Baton Rouge	36	HH
204.02	West Baton Rouge	33	LH
205	Lafourche	1	LL
205.02	St. Martin	1	LL
205.08	Jefferson	1	LL
205.14	Jefferson	1	LL
206	Jefferson	3	LL
207.03	Lafourche	1	LL
21.02	Lafayette	1	LL
216.02	Lafourche	1	LL
217	Lafourche	1	LL
22	East Baton Rouge	37	HH
22	Lafayette	1	LL
221.02	Jefferson	1	LL
23	East Baton Rouge	8	LH
231	Jefferson	1	LL
24	Calcasieu	1	LL
24	East Baton Rouge	71	HH
24.01	Orleans	1	LL
246	Jefferson	1	LL
25	East Baton Rouge	66	HH
251.02	Jefferson	2	LL
252.02	Jefferson	1	LL
26.01	East Baton Rouge	14	LH

(Table cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Result
26.02	East Baton Rouge	78	HH
261	Jefferson	1	LL
266	Jefferson	1	LL
27	Calcasieu	1	LL
27	East Baton Rouge	49	HH
27	Orleans	1	LL
270	Jefferson	1	LL
276.01	Jefferson	2	LL
277.01	Jefferson	2	LL
278.07	Jefferson	1	LL
28.01	East Baton Rouge	5	LH
28.02	East Baton Rouge	4	LH
3	East Baton Rouge	123	HH
30	East Baton Rouge	95	HH
30	Orleans	1	LL
301	Avoyelles	1	LL
301	Iberia	1	LL
301.01	Ascension	49	HH
301.02	Ascension	25	LH
301.03	Ascension	32	LH
302.03	Ascension	47	HH
302.04	Ascension	79	HH
302.05	Ascension	59	HH
302.06	Ascension	83	HH
303	Ascension	138	HH
303	Avoyelles	1	LL
303.01	Iberia	1	LL
304	Avoyelles	2	LL
304.01	Ascension	87	HH
304.02	Ascension	125	HH
305	Ascension	66	HH
306	Ascension	80	HH
307	Avoyelles	1	LL
309	Ascension	18	LH
31.01	Calcasieu	2	LL
31.01	East Baton Rouge	83	HH
31.03	East Baton Rouge	31	LH
310	Ascension	28	LH
311	Iberia	1	LL
32.01	East Baton Rouge	53	HH
32.02	East Baton Rouge	26	LH
33	East Baton Rouge	86	HH

(Table cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Result
33.03	Orleans	1	LL
34	East Baton Rouge	197	HH
35.01	East Baton Rouge	31	LH
35.04	East Baton Rouge	140	HH
35.05	East Baton Rouge	117	HH
35.06	East Baton Rouge	82	HH
35.07	East Baton Rouge	65	HH
36.01	East Baton Rouge	45	HH
36.03	East Baton Rouge	43	HH
36.04	East Baton Rouge	137	HH
37.01	East Baton Rouge	46	HH
37.01	Orleans	1	LL
37.02	East Baton Rouge	75	HH
37.02	Orleans	1	LL
37.03	East Baton Rouge	61	HH
38.01	East Baton Rouge	128	HH
38.02	East Baton Rouge	43	HH
38.04	East Baton Rouge	23	LH
38.05	East Baton Rouge	86	HH
39.04	East Baton Rouge	63	HH
39.06	East Baton Rouge	43	HH
39.07	East Baton Rouge	36	HH
39.08	East Baton Rouge	31	LH
39.09	East Baton Rouge	51	HH
39.1	East Baton Rouge	49	HH
4	East Baton Rouge	90	HH
4.02	Ouachita	1	LL
40.05	East Baton Rouge	86	HH
40.06	East Baton Rouge	34	HH
40.09	East Baton Rouge	57	HH
40.1	East Baton Rouge	24	LH
40.11	East Baton Rouge	141	HH
40.13	East Baton Rouge	79	HH
40.14	East Baton Rouge	55	HH
40.15	East Baton Rouge	167	HH
40.16	East Baton Rouge	49	HH
401	Livingston	74	HH
401.02	St. Tammany	2	LL
402	St. Mary	1	LL
402.01	Livingston	411	HH
402.02	Livingston	143	HH

(Table cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Result
403.01	Livingston	192	HH
403.03	Livingston	124	HH
403.04	Livingston	155	HH
403.05	St. Tammany	1	LL
404	St. Tammany	1	LL
404.01	Livingston	256	HH
404.02	Livingston	280	HH
405	Livingston	250	HH
405.02	St. Tammany	1	LL
406	Livingston	160	HH
407	Livingston	71	HH
407	St. James	2	LL
407.05	St. Tammany	1	LL
408.01	St. Tammany	2	LL
408.02	Livingston	114	HH
408.03	St. Tammany	1	LL
408.04	Livingston	673	HH
408.05	Livingston	103	HH
408.06	Livingston	162	HH
409.01	Livingston	48	HH
409.02	Livingston	38	HH
411.03	St. Tammany	1	LL
412.12	St. Tammany	1	LL
414	St. Mary	2	LL
42.01	East Baton Rouge	61	HH
42.03	East Baton Rouge	39	HH
42.04	East Baton Rouge	51	HH
42.05	East Baton Rouge	64	HH
43.01	East Baton Rouge	31	LH
43.02	East Baton Rouge	19	LH
44.01	East Baton Rouge	22	LH
44.02	East Baton Rouge	29	LH
44.03	East Baton Rouge	21	LH
45	Orleans	1	LL
45.03	East Baton Rouge	76	HH
45.04	East Baton Rouge	36	HH
45.05	East Baton Rouge	69	HH
45.07	East Baton Rouge	17	LH
45.08	East Baton Rouge	16	LH
45.09	East Baton Rouge	53	HH
45.1	East Baton Rouge	41	HH

(Table cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Result
46.02	East Baton Rouge	33	LH
46.03	East Baton Rouge	15	LH
46.04	East Baton Rouge	25	LH
47	East Baton Rouge	27	LH
48	East Baton Rouge	20	LH
49	East Baton Rouge	18	LH
5	Concordia	1	LL
5	East Baton Rouge	132	HH
5	Jefferson Davis	1	LL
5	Terrebonne	2	LL
50	East Baton Rouge	4	LH
505	Assumption	2	LL
51	East Baton Rouge	76	HH
52	East Baton Rouge	65	HH
53	East Baton Rouge	202	HH
6	Calcasieu	1	LL
6.01	East Baton Rouge	71	HH
6.02	East Baton Rouge	116	HH
6.07	Orleans	1	LL
6.11	Orleans	1	LL
6.12	Orleans	1	LL
6.13	Orleans	1	LL
60	Orleans	2	LL
623.01	St. Charles	1	LL
629	St. Charles	1	LL
63	Orleans	1	LL
7.01	East Baton Rouge	63	HH
7.02	East Baton Rouge	74	HH
702	St. John the Baptist	2	LL
703	St. John the Baptist	1	LL
705	St. John the Baptist	5	LL
707	St. John the Baptist	1	LL
708	St. John the Baptist	1	LL
711	St. John the Baptist	1	LL
85	Orleans	1	LL
9	East Baton Rouge	108	HH
9502	Washington	1	LL
9503	Washington	2	LL
9506	Morehouse	2	LL
9506	Washington	1	LL
9507	Vermilion	1	LL

(Table cont'd)

Census Tract ID	Parish	ED High Utilizers Count	Result
9512	St. Helena	31	LH
9515.01	East Feliciana	4	LH
9515.02	East Feliciana	28	LH
9516	East Feliciana	14	LH
9518	West Feliciana	13	LH
9520	Pointe Coupee	6	LL
9521	Pointe Coupee	23	LH
9522	Pointe Coupee	16	LH
9526	Iberville	36	HH
9527	Iberville	35	HH
9529	Iberville	26	LH
9530	Iberville	1	LH
9531.01	Iberville	20	LH
9531.02	Iberville	15	LH
9532	Iberville	78	HH
9532	Tangipahoa	1	LL
9533	Tangipahoa	2	LL
9535	Tangipahoa	3	LL
9536	Tangipahoa	1	LH
9546	Tangipahoa	2	LL
9548	Tangipahoa	3	LL
9601	St. Landry	3	LL
9602	Acadia	1	LL
9604	St. Landry	2	LL
9605	Acadia	2	LL
9607	Acadia	1	LL
9610	Acadia	1	LL
9610	St. Landry	1	LL
9613	St. Landry	1	LL
9616	St. Landry	2	LL
9618	St. Landry	1	LL
9705	Richland	3	LL

Appendix R. Spearman Correlation Scores between independent variables of Non-Institutionalized COVID Positive Patients Data

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	BMI	Age	LOS	Sex	CPMI
BMI	***	-0.14	0.03	-0.16	0.11
Age	-0.14	***	0.11	0.06	-0.85
LOS	0.03	0.11	***	0.04	-0.17
Living_Status	-0.01	0.19	0.18	0.07	-0.24
Sex	-0.16	0.06	0.04	***	-0.09
COVID_results	0.14	-0.10	0.19	-0.02	0.08
CPMI	0.11	-0.85	-0.17	-0.09	***
E_TOTPOP	-0.01	-0.03	-0.05	0.01	0.03
EP_POV	0.04	-0.01	0.09	-0.03	-0.01
EP_UNEMP	0.06	-0.01	0.06	-0.04	-0.03
EP_PCI	-0.06	0.01	-0.08	0.03	0.01
EP_NOHSDP	0.06	0.03	0.06	-0.01	-0.05
EP_AGE65	-0.04	0.10	0.03	-0.01	-0.08
EP_AGE17	0.04	-0.03	0.01	-0.01	0.03
EP_DISABL	0.03	0.05	0.04	-0.02	-0.08
EP_SNGPNT	0.04	-0.03	0.07	-0.03	0.01
EP_MINRTY	0.07	-0.02	0.06	-0.04	-0.02
EP_LIMENG	-0.01	-0.02	-0.01	0.01	-0.01
EP_MUNIT	-0.03	-0.01	0.00	0.00	-0.01
EP_MOBILE	0.02	-0.01	0.02	0.03	0.02
EP_CROWD	0.02	0.01	0.02	-0.02	-0.02
EP_NOVEH	0.05	0.00	0.05	-0.03	-0.03
EP_GROUPQ	0.02	0.02	0.06	-0.01	-0.02
RPL_THEME1	0.06	0.00	0.08	-0.03	-0.03
RPL_THEME2	0.04	0.03	0.06	-0.03	-0.05
RPL_THEME3	0.04	-0.03	0.04	-0.02	-0.02
RPL_THEME4	0.04	0.01	0.06	-0.01	-0.03
RPL_THEMES	0.06	0.01	0.08	-0.03	-0.04
EP_UNINSUR	0.04	-0.02	0.04	-0.02	0.02
EP_OCC_MBSA	-0.07	0.00	-0.06	0.01	0.02
EP_OCC_SER	0.04	-0.03	0.05	-0.02	-0.01
EP_OCC_SAL_OFF	0.00	0.00	0.01	0.01	0.01
EP_OCC_NRCM	0.01	0.03	0.03	0.03	-0.03
EP_OCC_PTMM	0.07	-0.01	0.02	-0.03	0.00
Auto_immun	0.01	0.05	0.02	-0.08	-0.07
CEVD	-0.07	0.14	0.05	0.04	-0.17

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	BMI	Age	LOS	Sex	CPMI
CKD	-0.03	0.19	0.15	0.06	-0.29
CLD	0.02	0.04	0.01	0.00	-0.07
CVD	-0.02	0.24	0.13	0.08	-0.31
Diabetes	0.09	0.23	0.11	0.02	-0.37
ESRD	-0.04	0.04	0.09	0.04	-0.12
Hep_B	-0.01	-0.01	0.00	0.02	0.00
Hep_C	-0.02	0.04	0.00	0.01	-0.05
HIV	0.00	-0.03	0.01	0.01	-0.02
HTN	0.07	0.37	0.11	0.01	-0.42
Immunocompromised	0.01	0.01	0.02	-0.02	-0.04
Obesity	0.37	0.02	0.18	-0.05	-0.10
overweight	-0.05	0.01	0.05	-0.01	-0.03
PVD	-0.05	0.10	0.05	0.02	-0.14
Respiratory	0.00	0.09	0.03	-0.02	-0.20

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	E_Tot Pop	Ep_Pov	EP_Unemp	EP_PCI	EP_NoHsdp	EP_Age65	EP_Age17
BMI	-0.01	0.04	0.06	-0.06	0.06	-0.04	0.04
Age	-0.03	-0.01	-0.01	0.01	0.03	0.10	-0.03
LOS	-0.05	0.09	0.06	-0.08	0.06	0.03	0.01
Living_Status	-0.04	0.04	0.03	-0.05	0.04	0.01	0.00
Sex	0.01	-0.03	-0.04	0.03	-0.01	-0.01	-0.01
COVID_results	-0.01	0.08	0.10	-0.08	0.06	-0.05	0.05
CPMI	0.03	-0.01	-0.03	0.01	-0.05	-0.08	0.03
E_TOTPOP	***	-0.53	-0.31	0.50	-0.43	-0.35	0.22
EP_POV	-0.53	***	0.63	-0.84	0.71	-0.05	0.04
EP_UNEMP	-0.31	0.63	***	-0.61	0.55	-0.07	0.14
EP_PCI	0.50	-0.84	-0.61	***	-0.75	0.11	-0.18
EP_NOHSDP	-0.43	0.71	0.55	-0.75	***	0.04	0.09
EP_AGE65	-0.35	-0.05	-0.07	0.11	0.04	***	-0.50
EP_AGE17	0.22	0.04	0.14	-0.18	0.09	-0.50	***
EP_DISABL	-0.47	0.49	0.41	-0.50	0.61	0.29	-0.14
EP_SNGPNT	-0.14	0.58	0.43	-0.61	0.47	-0.30	0.43
EP_MINRTY	-0.35	0.63	0.55	-0.64	0.44	-0.12	0.12
EP_LIMENG	0.23	-0.12	-0.14	0.16	-0.07	-0.12	-0.10
EP_MUNIT	0.00	0.10	-0.01	0.02	-0.17	0.00	-0.18
EP_MOBILE	0.10	-0.09	-0.06	0.00	0.28	0.03	0.12
EP_CROWD	-0.05	0.40	0.28	-0.41	0.38	-0.26	0.32
EP_NOVEH	-0.52	0.74	0.53	-0.64	0.63	0.09	-0.02
EP_GROUPQ	-0.22	0.28	0.14	-0.26	0.23	0.16	-0.20
RPL_THEME1	-0.51	0.91	0.80	-0.91	0.86	-0.05	0.12
RPL_THEME2	-0.39	0.57	0.47	-0.61	0.61	0.24	0.35
RPL_THEME3	-0.04	0.32	0.24	-0.29	0.22	-0.17	0.00
RPL_THEME4	-0.26	0.57	0.34	-0.52	0.52	-0.01	0.02
RPL_THEMES	-0.46	0.84	0.66	-0.83	0.79	0.01	0.14
EP_UNINSUR	-0.45	0.64	0.43	-0.65	0.59	-0.03	-0.06
EP_OCC_MBSA	0.39	-0.66	-0.51	0.78	-0.74	0.04	-0.12
EP_OCC_SER	-0.37	0.70	0.52	-0.73	0.52	-0.11	0.05
EP_OCC_SAL_OFF	0.00	-0.03	-0.03	0.00	-0.14	-0.07	0.07
EP_OCC_NRCM	0.02	0.01	-0.02	-0.07	0.34	0.06	0.02
EP_OCC_PTMM	-0.20	0.27	0.30	-0.42	0.47	0.04	0.12
Auto_immun	0.02	-0.03	-0.01	0.02	-0.03	0.01	-0.01
CEVD	0.01	0.01	0.00	-0.01	0.01	-0.03	-0.01
CKD	0.00	0.03	0.03	-0.03	0.04	0.03	-0.02
CLD	-0.01	-0.02	0.01	0.01	-0.01	0.02	-0.03
CVD	-0.01	0.03	0.04	-0.02	0.04	0.02	0.00

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	E_Tot Pop	Ep_Pov	EP_Unemp	EP_PCI	EP_NoHsdp	EP_Age65	EP_Age17
Diabetes	-0.02	0.05	0.06	-0.06	0.06	0.01	-0.01
ESRD	-0.05	0.07	0.06	-0.06	0.06	0.02	-0.02
Hep_B	0.01	0.01	0.01	-0.01	0.02	0.02	-0.02
Hep_C	-0.01	0.01	0.02	-0.02	0.04	0.00	-0.01
HIV	-0.03	0.03	0.05	-0.04	0.04	0.02	-0.02
HTN	-0.01	0.03	0.05	-0.03	0.04	0.04	-0.02
Immunocompromised	0.03	-0.04	-0.02	0.04	-0.04	-0.01	0.00
Obesity	0.04	0.04	0.08	-0.04	0.05	-0.04	0.01
overweight	0.02	-0.03	-0.03	0.05	-0.03	0.04	-0.03
PVD	0.00	0.01	0.01	0.00	0.02	0.01	0.01
Respiratory	0.00	0.02	0.01	0.00	0.03	0.01	0.01

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	Ep_Disabl	EP_Sngpnt	EP_Minrty	EP_LimEng	EP_Munit	EP_Mobile	EP_Crowd
BMI	0.03	0.04	0.07	-0.01	-0.03	0.02	0.02
Age	0.05	-0.03	-0.02	-0.02	-0.01	-0.01	0.01
LOS	0.04	0.07	0.06	-0.01	0.00	0.02	0.02
Living_Status	0.03	0.03	0.06	0.02	-0.01	-0.02	0.03
Sex	-0.02	-0.03	-0.04	0.01	0.00	0.03	-0.02
COVID_results	0.03	0.10	0.19	0.05	0.03	-0.08	0.03
CPMI	-0.08	0.01	-0.02	-0.01	-0.01	0.02	-0.02
E_TOTPOP	-0.47	-0.14	-0.35	0.23	0.00	0.10	-0.05
EP_POV	0.49	0.58	0.63	-0.12	0.10	-0.09	0.40
EP_UNEMP	0.41	0.43	0.55	-0.14	-0.01	-0.06	0.28
EP_PCI	-0.50	-0.61	-0.64	0.16	0.02	0.00	-0.41
EP_NOHSDP	0.61	0.47	0.44	-0.07	-0.17	0.28	0.38
EP_AGE65	0.29	-0.30	-0.12	-0.12	0.00	0.03	-0.26
EP_AGE17	-0.14	0.43	0.12	-0.10	-0.18	0.12	0.32
EP_DISABL	***	0.19	0.34	-0.09	-0.12	0.15	0.19
EP_SNGPNT	0.19	***	0.58	-0.05	0.12	-0.09	0.37
EP_MINRTY	0.34	0.58	***	0.06	0.28	-0.42	0.33
EP_LIMENG	-0.09	-0.05	0.06	***	0.32	-0.06	0.07
EP_MUNIT	-0.12	0.12	0.28	0.32	***	-0.62	0.02
EP_MOBILE	0.15	-0.09	-0.42	-0.06	-0.62	***	0.07
EP_CROWD	0.19	0.37	0.33	0.07	0.02	0.07	***
EP_NOVEH	0.50	0.47	0.63	-0.09	0.21	-0.17	0.30
EP_GROUPQ	0.10	0.22	0.27	-0.01	0.31	-0.20	0.02
RPL_THEME1	0.58	0.59	0.64	-0.14	-0.03	0.04	0.42
RPL_THEME2	0.69	0.66	0.46	-0.16	-0.09	0.11	0.32
RPL_THEME3	0.15	0.33	0.67	0.77	0.40	-0.30	0.27
RPL_THEME4	0.29	0.45	0.46	0.11	0.44	-0.04	0.56
RPL_THEMES	0.58	0.68	0.73	0.08	0.21	-0.04	0.53
EP_UNINSUR	0.35	0.44	0.48	0.06	0.13	0.04	0.25
EP_OCC_MBSA	-0.52	-0.52	-0.53	0.05	0.08	-0.16	-0.40
EP_OCC_SER	0.38	0.52	0.66	-0.02	0.21	-0.23	0.32
EP_OCC_SAL_OFF	-0.08	0.05	-0.01	-0.12	-0.05	-0.07	0.02
EP_OCC_NRCM	0.19	-0.05	-0.22	0.13	-0.29	0.57	0.20
EP_OCC_PTMM	0.41	0.31	0.27	-0.02	-0.21	0.31	0.13
Auto_immun	0.00	-0.01	0.00	-0.01	0.00	0.00	-0.01
CEVD	0.00	0.02	0.02	0.01	0.04	-0.02	0.04
CKD	0.05	0.00	0.04	-0.01	-0.01	-0.01	0.02
CLD	0.01	-0.02	0.00	0.03	0.04	-0.02	-0.01
CVD	0.04	0.03	0.02	-0.01	0.01	0.01	0.04

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	Ep_Disabl	EP_Sngpnt	EP_Minrty	EP_LimEng	EP_Munit	EP_Mobile	EP_Crowd
Diabetes	0.07	0.03	0.08	0.01	0.01	-0.02	0.05
ESRD	0.06	0.03	0.06	-0.02	0.01	-0.01	0.04
Hep_B	0.04	0.02	0.04	0.03	0.04	-0.04	-0.02
Hep_C	0.04	0.02	0.03	0.01	0.03	-0.03	-0.01
HIV	0.06	0.02	0.05	0.02	0.04	-0.05	0.00
HTN	0.08	0.03	0.08	-0.03	0.03	-0.05	0.02
Immunocompromised	-0.03	-0.03	-0.01	0.03	0.01	-0.02	-0.02
Obesity	0.05	0.06	0.10	0.03	0.02	-0.04	0.04
overweight	-0.02	-0.03	-0.01	-0.02	0.01	-0.03	-0.02
PVD	0.01	0.01	0.00	0.01	-0.01	-0.01	0.01
Respiratory	0.02	0.04	0.00	0.01	0.00	0.02	0.03

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	EP_Noveh	EP_GroupQ	RPL_Theme1	RPL_Theme2	RPL_Theme3	RPL_Theme4	RPL_Themes
BMI	0.05	0.02	0.06	0.04	0.04	0.04	0.06
Age	0.00	0.02	0.00	0.03	-0.03	0.01	0.01
LOS	0.05	0.06	0.08	0.06	0.04	0.06	0.08
Living_Status	0.04	-0.01	0.05	0.03	0.06	0.01	0.05
Sex	-0.03	-0.01	-0.03	-0.03	-0.02	-0.01	-0.03
COVID_results	0.09	0.00	0.09	0.07	0.17	0.03	0.11
CPMI	-0.03	-0.02	-0.03	-0.05	-0.02	-0.03	-0.04
E_TOTPOP	-0.52	-0.22	-0.51	-0.39	-0.04	-0.26	-0.46
EP_POV	0.74	0.28	0.91	0.57	0.32	0.57	0.84
EP_UNEMP	0.53	0.14	0.80	0.47	0.24	0.34	0.66
EP_PCI	-0.64	-0.26	-0.91	-0.61	-0.29	-0.52	-0.83
EP_NOHSDP	0.63	0.23	0.86	0.61	0.22	0.52	0.79
EP_AGE65	0.09	0.16	-0.05	0.24	-0.17	-0.01	0.01
EP_AGE17	-0.02	-0.20	0.12	0.35	0.00	0.02	0.14
EP_DISABL	0.50	0.10	0.58	0.69	0.15	0.29	0.58
EP_SNGPNT	0.47	0.22	0.59	0.66	0.33	0.45	0.68
EP_MINRTY	0.63	0.27	0.64	0.46	0.67	0.46	0.73
EP_LIMENG	-0.09	-0.01	-0.14	-0.16	0.77	0.11	0.08
EP_MUNIT	0.21	0.31	-0.03	-0.09	0.40	0.44	0.21
EP_MOBILE	-0.17	-0.20	0.04	0.11	-0.30	-0.04	-0.04
EP_CROWD	0.30	0.02	0.42	0.32	0.27	0.56	0.53
EP_NOVEH	***	0.32	0.72	0.51	0.33	0.65	0.78
EP_GROUPQ	0.32	***	0.27	0.15	0.14	0.68	0.43
RPL_THEME1	0.72	0.27	***	0.65	0.30	0.56	0.89
RPL_THEME2	0.51	0.15	0.65	***	0.17	0.38	0.73
RPL_THEME3	0.33	0.14	0.30	0.17	***	0.36	0.52
RPL_THEME4	0.65	0.68	0.56	0.38	0.36	***	0.79
RPL_THEMES	0.78	0.43	0.89	0.73	0.52	0.79	***
EP_UNINSUR	0.55	0.24	0.66	0.37	0.34	0.48	0.66
EP_OCC_MBSA	-0.55	-0.14	-0.76	-0.56	-0.29	-0.45	-0.72
EP_OCC_SER	0.57	0.18	0.71	0.44	0.40	0.42	0.68
EP_OCC_SAL_OFF	-0.04	-0.04	-0.06	-0.04	-0.07	-0.08	-0.08
EP_OCC_NRCM	-0.03	-0.07	0.12	0.14	-0.06	0.13	0.11
EP_OCC_PTMM	0.24	0.08	0.42	0.43	0.15	0.21	0.41
Auto_immun	-0.03	0.01	-0.02	-0.01	-0.01	-0.01	-0.02
CEVD	0.01	0.00	0.01	-0.01	0.02	0.03	0.02
CKD	0.02	0.01	0.04	0.03	0.02	0.01	0.04
CLD	0.00	0.00	-0.01	-0.01	0.02	0.01	0.00

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	EP_Noveh	EP_GroupQ	RPL_Theme1	RPL_Theme2	RPL_Theme3	RPL_Theme4	RPL_Themes
CVD	0.03	0.01	0.04	0.04	0.01	0.04	0.04
Diabetes	0.05	0.01	0.07	0.05	0.06	0.04	0.07
ESRD	0.07	0.03	0.07	0.04	0.03	0.06	0.07
Hep_B	0.03	0.01	0.01	0.02	0.04	0.02	0.03
Hep_C	0.02	-0.02	0.03	0.02	0.02	0.00	0.03
HIV	0.04	0.02	0.05	0.04	0.05	0.02	0.05
HTN	0.05	0.00	0.04	0.06	0.02	0.02	0.05
Immunocompromised	-0.05	0.00	-0.04	-0.03	0.01	-0.02	-0.03
Obesity	0.04	0.00	0.06	0.04	0.08	0.03	0.06
overweight	-0.02	0.01	-0.04	-0.03	-0.02	-0.02	-0.04
PVD	0.01	-0.01	0.02	0.02	0.01	0.00	0.01
Respiratory	0.01	0.00	0.02	0.04	0.01	0.02	0.02

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	EP_Uninsur	EP_MBSA	EP_SER	EP_SAL	EP_NRCM	EP_PTMM	Auto_Immune
BMI	0.04	-0.07	0.04	0.00	0.01	0.07	0.01
Age	-0.02	0.00	-0.03	0.00	0.03	-0.01	0.05
LOS	0.04	-0.06	0.05	0.01	0.03	0.02	0.02
Living_Status	0.03	-0.06	0.05	-0.02	0.02	0.03	0.01
Sex	-0.02	0.01	-0.02	0.01	0.03	-0.03	-0.08
COVID_results	0.07	-0.08	0.10	-0.02	-0.03	0.06	-0.05
CPMI	0.02	0.02	-0.01	0.01	-0.03	0.00	-0.07
E_TOTPOP	-0.45	0.39	-0.37	0.00	0.02	-0.20	0.02
EP_POV	0.64	-0.66	0.70	-0.03	0.01	0.27	-0.03
EP_UNEMP	0.43	-0.51	0.52	-0.03	-0.02	0.30	-0.01
EP_PCI	-0.65	0.78	-0.73	0.00	-0.07	-0.42	0.02
EP_NOHSDP	0.59	-0.74	0.52	-0.14	0.34	0.47	-0.03
EP_AGE65	-0.03	0.04	-0.11	-0.07	0.06	0.04	0.01
EP_AGE17	-0.06	-0.12	0.05	0.07	0.02	0.12	-0.01
EP_DISABL	0.35	-0.52	0.38	-0.08	0.19	0.41	0.00
EP_SNGPNT	0.44	-0.52	0.52	0.05	-0.05	0.31	-0.01
EP_MINRTY	0.48	-0.53	0.66	-0.01	-0.22	0.27	0.00
EP_LIMENG	0.06	0.05	-0.02	-0.12	0.13	-0.02	-0.01
EP_MUNIT	0.13	0.08	0.21	-0.05	-0.29	-0.21	0.00
EP_MOBILE	0.04	-0.16	-0.23	-0.07	0.57	0.31	0.00
EP_CROWD	0.25	-0.40	0.32	0.02	0.20	0.13	-0.01
EP_NOVEH	0.55	-0.55	0.57	-0.04	-0.03	0.24	-0.03
EP_GROUPQ	0.24	-0.14	0.18	-0.04	-0.07	0.08	0.01
RPL_THEME1	0.66	-0.76	0.71	-0.06	0.12	0.42	-0.02
RPL_THEME2	0.37	-0.56	0.44	-0.04	0.14	0.43	-0.01
RPL_THEME3	0.34	-0.29	0.40	-0.07	-0.06	0.15	-0.01
RPL_THEME4	0.48	-0.45	0.42	-0.08	0.13	0.21	-0.01
RPL_THEMES	0.66	-0.72	0.68	-0.08	0.11	0.41	-0.02
EP_UNINSUR	***	-0.63	0.55	-0.08	0.20	0.31	-0.03
EP_OCC_MBSA	-0.63	***	-0.68	-0.06	-0.30	-0.63	0.02
EP_OCC_SER	0.55	-0.68	***	-0.13	-0.14	0.22	-0.01
EP_OCC_SAL_OFF	-0.08	-0.06	-0.13	***	-0.26	-0.16	-0.01
EP_OCC_NRCM	0.20	-0.30	-0.14	-0.26	***	0.22	0.01
EP_OCC_PTMM	0.31	-0.63	0.22	-0.16	0.22	***	0.00
Auto_immun	-0.03	0.02	-0.01	-0.01	0.01	0.00	***
CEVD	0.01	-0.02	0.02	-0.02	0.01	0.00	0.02
CKD	0.01	-0.05	0.05	-0.02	0.02	0.02	0.01
CLD	0.03	-0.02	0.01	-0.01	0.01	0.03	0.05
CVD	0.01	-0.05	0.03	-0.01	0.02	0.02	0.01

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	EP_Uninsur	EP_MBSA	EP_SER	EP_SAL	EP_NRCM	EP_PTMM	Auto_Immune
Diabetes	0.04	-0.07	0.08	-0.03	0.02	0.03	-0.01
ESRD	0.04	-0.06	0.05	0.03	0.01	0.00	0.00
Hep_B	0.00	-0.04	0.03	-0.01	0.01	0.04	-0.02
Hep_C	0.01	-0.02	0.02	-0.03	0.01	0.02	0.01
HIV	0.02	-0.05	0.05	-0.03	0.00	0.03	-0.02
HTN	0.01	-0.06	0.04	0.00	0.01	0.04	0.06
Immunocompromised	-0.04	0.02	0.01	-0.02	-0.01	0.00	0.14
Obesity	0.01	-0.09	0.08	0.01	0.01	0.06	0.04
overweight	-0.04	0.04	-0.02	-0.01	-0.02	-0.03	0.00
PVD	0.01	0.01	0.01	0.00	-0.01	-0.01	0.00
Respiratory	0.01	-0.02	0.01	-0.02	0.04	0.02	0.03

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	CEVD	CKD	CLD	CVD	Diabetes	ESRD	Hep_B
BMI	-0.07	-0.03	0.02	-0.02	0.09	-0.04	-0.01
Age	0.14	0.19	0.04	0.24	0.23	0.04	-0.01
LOS	0.05	0.15	0.01	0.13	0.11	0.09	0.00
Living_Status	0.11	0.16	0.02	0.13	0.14	0.11	0.02
Sex	0.04	0.06	0.00	0.08	0.02	0.04	0.02
COVID_results	0.02	0.01	-0.01	-0.01	0.07	0.04	0.00
CPMI	-0.17	-0.29	-0.07	-0.31	-0.37	-0.12	0.00
E_TOTPOP	0.01	0.00	-0.01	-0.01	-0.02	-0.05	0.01
EP_POV	0.01	0.03	-0.02	0.03	0.05	0.07	0.01
EP_UNEMP	0.00	0.03	0.01	0.04	0.06	0.06	0.01
EP_PCI	-0.01	-0.03	0.01	-0.02	-0.06	-0.06	-0.01
EP_NOHSDP	0.01	0.04	-0.01	0.04	0.06	0.06	0.02
EP_AGE65	-0.03	0.03	0.02	0.02	0.01	0.02	0.02
EP_AGE17	-0.01	-0.02	-0.03	0.00	-0.01	-0.02	-0.02
EP_DISABL	0.00	0.05	0.01	0.04	0.07	0.06	0.04
EP_SNGPNT	0.02	0.00	-0.02	0.03	0.03	0.03	0.02
EP_MINRTY	0.02	0.04	0.00	0.02	0.08	0.06	0.04
EP_LIMENG	0.01	-0.01	0.03	-0.01	0.01	-0.02	0.03
EP_MUNIT	0.04	-0.01	0.04	0.01	0.01	0.01	0.04
EP_MOBILE	-0.02	-0.01	-0.02	0.01	-0.02	-0.01	-0.04
EP_CROWD	0.04	0.02	-0.01	0.04	0.05	0.04	-0.02
EP_NOVEH	0.01	0.02	0.00	0.03	0.05	0.07	0.03
EP_GROUPQ	0.00	0.01	0.00	0.01	0.01	0.03	0.01
RPL_THEME1	0.01	0.04	-0.01	0.04	0.07	0.07	0.01
RPL_THEME2	-0.01	0.03	-0.01	0.04	0.05	0.04	0.02
RPL_THEME3	0.02	0.02	0.02	0.01	0.06	0.03	0.04
RPL_THEME4	0.03	0.01	0.01	0.04	0.04	0.06	0.02
RPL_THEMES	0.02	0.04	0.00	0.04	0.07	0.07	0.03
EP_UNINSUR	0.01	0.01	0.03	0.01	0.04	0.04	0.00
EP_OCC_MBSA	-0.02	-0.05	-0.02	-0.05	-0.07	-0.06	-0.04
EP_OCC_SER	0.02	0.05	0.01	0.03	0.08	0.05	0.03
EP_OCC_SAL_OFF	-0.02	-0.02	-0.01	-0.01	-0.03	0.03	-0.01
EP_OCC_NRCM	0.01	0.02	0.01	0.02	0.02	0.01	0.01
EP_OCC_PTMM	0.00	0.02	0.03	0.02	0.03	0.00	0.04
Auto_immun	0.02	0.01	0.05	0.01	-0.01	0.00	-0.02
CEVD	***	0.16	0.00	0.18	0.13	0.05	0.00
CKD	0.16	***	0.03	0.35	0.28	0.31	0.04
CLD	0.00	0.03	***	0.04	0.06	0.01	0.07
CVD	0.18	0.35	0.04	***	0.24	0.15	0.00

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	CEVD	CKD	CLD	CVD	Diabetes	ESRD	Hep_B
Diabetes	0.13	0.28	0.06	0.24	***	0.12	0.01
ESRD	0.05	0.31	0.01	0.15	0.12	***	0.05
Hep_B	0.00	0.04	0.07	0.00	0.01	0.05	***
Hep_C	0.02	0.02	0.15	0.02	0.02	0.02	0.14
HIV	0.00	0.05	0.02	0.00	0.03	0.06	0.26
HTN	0.16	0.26	0.08	0.28	0.38	0.09	0.02
Immunocompromised	0.06	0.02	-0.01	0.03	0.02	-0.01	-0.01
Obesity	0.02	0.10	0.02	0.12	0.23	0.06	-0.01
overweight	0.00	0.01	-0.01	0.00	0.01	-0.01	-0.01
PVD	0.09	0.15	0.04	0.21	0.12	0.09	0.02
Respiratory	0.03	0.07	0.01	0.15	0.07	0.02	0.00

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	Hep_C	HIV	HTN	Immu_Com	Obesity	Overweight	PVD	Respiratory
BMI	-0.02	0.00	0.07	0.01	0.37	-0.05	-0.05	0.00
Age	0.04	-0.03	0.37	0.01	0.02	0.01	0.10	0.09
LOS	0.00	0.01	0.11	0.02	0.18	0.05	0.05	0.03
Living_Status	0.02	0.03	0.12	0.01	0.09	0.01	0.05	0.06
Sex	0.01	0.01	0.01	-0.02	-0.05	-0.01	0.02	-0.02
COVID_results	-0.04	0.00	-0.02	-0.01	0.19	0.03	0.00	-0.07
CPMI	-0.05	-0.02	-0.42	-0.04	-0.10	-0.03	-0.14	-0.20
E_TOTPOP	-0.01	-0.03	-0.01	0.03	0.04	0.02	0.00	0.00
EP_POV	0.01	0.03	0.03	-0.04	0.04	-0.03	0.01	0.02
EP_UNEMP	0.02	0.05	0.05	-0.02	0.08	-0.03	0.01	0.01
EP_PCI	-0.02	-0.04	-0.03	0.04	-0.04	0.05	0.00	0.00
EP_NOHSDP	0.04	0.04	0.04	-0.04	0.05	-0.03	0.02	0.03
EP_AGE65	0.00	0.02	0.04	-0.01	-0.04	0.04	0.01	0.01
EP_AGE17	-0.01	-0.02	-0.02	0.00	0.01	-0.03	0.01	0.01
EP_DISABL	0.04	0.06	0.08	-0.03	0.05	-0.02	0.01	0.02
EP_SNGPNT	0.02	0.02	0.03	-0.03	0.06	-0.03	0.01	0.04
EP_MINRTY	0.03	0.05	0.08	-0.01	0.10	-0.01	0.00	0.00
EP_LIMENG	0.01	0.02	-0.03	0.03	0.03	-0.02	0.01	0.01
EP_MUNIT	0.03	0.04	0.03	0.01	0.02	0.01	-0.01	0.00
EP_MOBILE	-0.03	-0.05	-0.05	-0.02	-0.04	-0.03	-0.01	0.02
EP_CROWD	-0.01	0.00	0.02	-0.02	0.04	-0.02	0.01	0.03
EP_NOVEH	0.02	0.04	0.05	-0.05	0.04	-0.02	0.01	0.01
EP_GROUPQ	-0.02	0.02	0.00	0.00	0.00	0.01	-0.01	0.00
RPL_THEME1	0.03	0.05	0.04	-0.04	0.06	-0.04	0.02	0.02
RPL_THEME2	0.02	0.04	0.06	-0.03	0.04	-0.03	0.02	0.04
RPL_THEME3	0.02	0.05	0.02	0.01	0.08	-0.02	0.01	0.01
RPL_THEME4	0.00	0.02	0.02	-0.02	0.03	-0.02	0.00	0.02
RPL_THEMES	0.03	0.05	0.05	-0.03	0.06	-0.04	0.01	0.02
EP_UNINSUR	0.01	0.02	0.01	-0.04	0.01	-0.04	0.01	0.01
EP_OCC_MBSA	-0.02	-0.05	-0.06	0.02	-0.09	0.04	0.01	-0.02
EP_OCC_SER	0.02	0.05	0.04	0.01	0.08	-0.02	0.01	0.01
EP_OCC_SAL_OFF	-0.03	-0.03	0.00	-0.02	0.01	-0.01	0.00	-0.02
EP_OCC_NRCM	0.01	0.00	0.01	-0.01	0.01	-0.02	-0.01	0.04
EP_OCC_PTMM	0.02	0.03	0.04	0.00	0.06	-0.03	-0.01	0.02
Auto_immun	0.01	-0.02	0.06	0.14	0.04	0.00	0.00	0.03
CEVD	0.02	0.00	0.16	0.06	0.02	0.00	0.09	0.03
CKD	0.02	0.05	0.26	0.02	0.10	0.01	0.15	0.07

(Table cont'd)

LEGEND

(+/-)0.1-039, Weak Correlation ; (+/-)0.4-0.69, Medium Correlation; (+/-)0.7-1, Strong Correlation

Spearman (r_s)	Hep_C	HIV	HTN	Immu_Com	Obesity	Overweight	PVD	Respiratory
CLD	0.15	0.02	0.08	-0.01	0.02	-0.01	0.04	0.01
CVD	0.02	0.00	0.28	0.03	0.12	0.00	0.21	0.15
Diabetes	0.02	0.03	0.38	0.02	0.23	0.01	0.12	0.07
ESRD	0.02	0.06	0.09	-0.01	0.06	-0.01	0.09	0.02
Hep_B	0.14	0.26	0.02	-0.01	-0.01	-0.01	0.02	0.00
Hep_C	***	0.09	0.07	-0.01	-0.01	-0.01	0.05	0.02
HIV	0.09	***	0.03	-0.01	0.01	-0.01	0.01	0.03
HTN	0.07	0.03	***	0.01	0.24	0.02	0.12	0.13
Immunocompromised	-0.01	-0.01	0.01	***	0.03	-0.01	-0.01	0.02
Obesity	-0.01	0.01	0.24	0.03	***	-0.03	0.01	0.09
overweight	-0.01	-0.01	0.02	-0.01	-0.03	***	0.00	-0.01
PVD	0.05	0.01	0.12	-0.01	0.01	0.00	***	0.07
Respiratory	0.02	0.03	0.13	0.02	0.09	-0.01	0.07	***

(Table cont'd)

Appendix S. Summary of COVID Positive Patients Based on Demographics

Study Characteristics	Positive %	Alive %	Dead %	Male %	Female %	Average CPMI	Average BMI	Average Age
Race_Ethnicity								
White or Caucasian	32.99	29.08	3.92	15.05	17.95	45.46	22.48	61.93
Black or African American	56.31	50.76	5.55	22.34	33.97	59.00	28.59	53.39
Asian	0.58	0.51	0.07	0.29	0.29	62.15	17.43	50.16
Other	1.27	1.12	0.15	0.58	0.69	69.38	16.99	54.17
Unknown	5.22	4.97	0.25	2.50	2.72	69.56	8.83	53.23
Hispanic	3.63	3.59	0.04	1.81	1.81	74.11	23.95	42.49
Patient Type								
ED	21.25	20.70	0.54	8.12	13.13	88.52	31.75	44.68
Admit	32.67	24.44	8.23	16.39	16.28	49.45	31.66	64.85
Outpatient /UC	46.08	44.89	1.20	18.06	28.03	45.17	17.52	54.50
Smoking Status								
Never	62.94	57.61	5.33	22.19	40.75	57.50	27.22	53.67
Former	20.56	16.79	3.77	11.17	9.39	40.54	28.01	65.49
Current	7.65	6.96	0.69	4.79	2.86	56.92	24.80	53.26
Unknown	8.85	8.67	0.18	4.42	4.42	78.00	4.21	50.57
Age Category								
11 to 20	2.83	2.79	0.04	1.16	1.67	85.36	19.89	19.54
21 to 30	11.24	11.17	0.07	4.24	7.00	75.73	23.72	26.39
31 to 40	11.71	11.57	0.15	4.39	7.32	76.60	26.68	35.92
41 to 50	14.10	13.67	0.44	5.66	8.45	68.14	27.61	46.35
51 to 60	18.02	16.46	1.56	7.47	10.55	66.66	26.68	56.28
61 to 70	18.35	15.99	2.36	9.35	8.99	50.84	25.67	65.77
71 to 80	12.26	9.72	2.54	6.24	6.02	30.15	24.80	75.48
81 to 90	8.67	6.56	2.10	3.37	5.29	12.40	22.04	85.57
91 to 100	2.72	1.99	0.73	0.65	2.07	5.98	16.23	93.66
101 to 110	0.11	0.11	0.00	0.04	0.07	0.00	6.16	102.81
Institutionalized								
No	79.91	74.00	5.91	32.70	47.21	62.18	25.89	51.28
Yes	20.09	16.03	4.06	9.86	10.22	30.32	22.28	73.78

(Table Cont'd)

Study Characteristics	Positive %	Alive %	Dead %	Male %	Female %	Average CPMI	Average BMI	Average Age
Institution Group								
Noninstitutionalized	79.91	74.00	5.91	32.70	47.21	62.18	25.89	51.28
Nursing Home	17.91	14.14	3.77	8.12	9.79	25.93	21.35	75.55
Rehab/Mental Health	0.29	0.22	0.07	0.29	0.00	56.22	27.16	71.97
Prison/Jail	1.89	1.67	0.22	1.45	0.44	67.98	30.35	57.19
Gender								
Male	42.57	36.77	5.80	42.57	57.43	54.66	23.66	56.52
Female	57.43	53.26	4.17	0.00	0.00	56.61	26.28	55.26
Living Status								
Deceased	9.97	0.00	9.97	5.80	4.17	36.26	30.23	71.55
Alive	90.03	90.03	0.00	36.77	53.26	57.94	24.60	54.05

Appendix T. Summary of Positive Patients Based on Underlying Conditions

Study Characteristics	% HTN	% Diabetes	% CVD	% CKD	% CEVD	% Obesity	%Respiratory	% Other Problems
Race_Ethnicity								
White or Caucasian	19.07	9.17	6.85	4.97	3.70	6.74	4.97	4.53
Black or African American	34.63	21.14	10.04	8.96	6.02	16.82	6.67	10.37
Asian	0.15	0.11	0.04	0.04	0.11	0.04	0.07	0.04
Other	0.40	0.18	0.11	0.11	0.07	0.15	0.04	0.15
Unknown	1.34	0.47	0.36	0.36	0.18	0.44	0.29	0.36
Hispanic	0.76	0.65	0.18	0.25	0.07	0.80	0.15	0.15
Patient Type								
ED	8.77	4.68	1.81	1.12	0.98	3.81	1.78	1.49
Admit	25.49	16.42	10.99	9.57	6.42	11.39	6.67	9.39
Outpatient /UC	22.08	10.62	4.79	3.99	2.76	9.79	3.73	4.71
Smoking Status								
Never	34.88	19.33	8.92	8.12	4.89	17.48	6.16	8.09
Former	15.77	9.25	6.60	4.97	3.77	6.13	3.88	5.18
Current	4.64	2.72	1.78	1.34	1.16	1.23	1.96	2.18
Unknown	1.05	0.44	0.29	0.25	0.33	0.15	0.18	0.15
Living Status								
Deceased	8.16	5.62	4.06	3.95	2.43	3.26	2.32	3.01
Alive	48.19	26.11	13.52	10.73	7.72	21.72	9.86	12.58
Gender								
Male	24.66	14.03	8.74	7.76	5.40	8.56	5.00	7.47
Female	31.69	17.69	8.85	6.93	4.75	16.42	7.18	8.12
Age Category								
11 to 20	0.15	0.11	0.00	0.00	0.04	0.33	0.15	0.07
21 to 30	0.87	0.58	0.04	0.07	0.07	1.56	0.76	0.29
31 to 40	2.61	1.60	0.15	0.36	0.07	3.08	0.73	0.83
41 to 50	7.32	3.66	0.18	0.94	0.44	4.86	1.27	1.45
51 to 60	11.57	6.53	0.54	2.07	1.63	6.20	2.18	3.95
61 to 70	13.92	8.96	0.36	4.13	3.30	5.29	2.68	3.81
71 to 80	10.48	5.98	0.36	3.41	2.54	2.65	2.43	3.23
81 to 90	7.11	3.44	0.11	2.72	1.74	0.91	1.56	1.45
91 to 100	2.21	0.87	0.04	0.94	0.29	0.11	0.44	0.47
101 to 110	0.11	0.00	0.00	0.04	0.04	0.00	0.00	0.04
Institutionalized								
No	38.98	22.63	10.26	8.56	4.79	21.14	8.12	10.66
Yes	17.37	9.10	7.32	6.13	5.37	3.84	4.06	4.93

(Table Cont'd)

Study Characteristics	% HTN	% Diabetes	% CVD	% CKD	% CEVD	% Obesity	%Respiratory	% Other Problems
Institution Group								
Non_Institutionalized	38.98	22.63	10.26	8.56	4.79	21.14	8.12	10.66
Nursing Home	15.95	8.52	6.93	5.77	5.11	3.19	3.70	4.50
Rehab/Mental Health	0.18	0.11	0.07	0.07	0.04	0.00	0.07	0.00
Prison/Jail	1.23	0.47	0.33	0.29	0.22	0.65	0.29	0.44

Appendix U. Summary of COVID Positive Patients Based on SVI and ACS Data

Study Characteristics	Average EP_POV	Average EP_UNEMP	Average EP_PCI	Average EP_NOHSDP	Average EP_AGE65	Average EP_AGE17
Race_Ethnicity						
White or Caucasian	13.97	5.18	29321.70	11.53	13.75	23.03
Black or AfricanAmerican	21.38	7.75	21712.82	15.06	12.43	22.17
Asian	12.95	4.83	31446.75	10.09	14.25	23.77
Other	16.16	5.89	23521.69	11.04	11.86	19.39
Unknown	18.86	5.87	25118.06	13.99	13.99	22.44
Hispanic	18.13	6.39	25051.42	13.27	12.13	23.10
Patient Type						
ED	21.17	7.49	21859.39	15.17	12.00	22.46
Admit	18.98	6.97	23137.72	14.98	12.55	21.74
Outpatient /UC	17.09	6.18	26903.71	12.11	13.65	23.00
Smoking Status						
Never	18.52	6.82	24581.58	13.62	12.69	22.73
Former	17.73	6.57	25077.44	13.65	13.27	21.94
Current	20.36	7.19	24940.34	13.18	13.12	21.97
Unknown	19.40	5.89	23345.71	14.81	13.80	22.33
Living Status						
Deceased	18.19	6.76	24076.03	13.98	13.45	21.70
Alive	18.62	6.71	24659.85	13.67	12.88	22.56
Gender						
Male	18.20	6.58	24947.05	13.98	12.98	22.14
Female	18.85	6.81	24345.63	13.49	12.91	22.73

(Table Cont'd)

Study Characteristics	Average EP_DISABL	Average EP_SNGPNT	Average EP_MINRTY	Average EP_LIMENG	Average EP_MUNIT	Average EP_MOBILE	Average EP_CROWD
Race_Ethnicity							
White or Caucasian	12.77	10.15	32.48	1.17	7.81	13.31	2.00
Black or African American	14.48	12.45	55.64	1.20	7.91	9.24	2.60
Asian	13.32	10.25	36.88	1.59	8.68	13.58	2.16
Other	11.47	9.46	36.46	0.68	8.68	7.44	2.34
Unknown	13.83	11.11	43.23	1.31	8.14	12.79	2.20
Hispanic	12.33	11.26	43.12	2.73	7.09	14.04	2.56
Patient Type							
ED	13.04	12.48	46.94	1.00	5.41	11.70	2.45
Admit	14.11	11.91	49.98	1.05	7.07	10.71	2.65
Outpatient /UC	13.85	10.81	43.93	1.50	9.58	10.77	2.14
Smoking Status							
Never	13.51	11.64	47.65	1.25	7.64	10.71	2.42
Former	14.14	11.35	44.98	1.19	8.21	10.97	2.19
Current	14.50	11.73	48.36	1.21	10.17	7.71	2.52
Unknown	14.04	10.96	40.74	1.39	6.78	15.39	2.40
Living Status							
Deceased	14.17	11.65	47.62	1.24	7.66	10.80	2.40
Alive	13.71	11.51	46.43	1.25	7.90	10.96	2.37
Gender							
Male	13.71	11.44	45.26	1.24	7.90	11.56	2.27
Female	13.80	11.59	47.50	1.25	7.86	10.50	2.46

(Table Cont'd)

Study Characteristics	Average EP_NOVEH	Average EP_GROUPQ	Average RPL THEME1	Average RPL THEME2	Average RPL THEME3	Average RPL THEME4	Average RPL THEMES
Race_Ethnicity							
White or Caucasian	5.93	2.81	0.31	0.40	0.43	0.45	0.36
Black or African American	9.70	3.75	0.49	0.51	0.54	0.53	0.53
Asian	6.81	3.29	0.25	0.40	0.51	0.45	0.33
Other	6.43	4.67	0.37	0.36	0.42	0.46	0.39
Unknown	8.78	4.64	0.41	0.46	0.48	0.50	0.45
Hispanic	6.47	2.80	0.41	0.38	0.54	0.48	0.43
Patient Type							
ED	8.73	3.42	0.48	0.48	0.47	0.52	0.51
Admit	8.12	5.38	0.45	0.47	0.51	0.53	0.49
Outpatient /UC	8.09	2.11	0.37	0.45	0.51	0.47	0.42
Smoking Status							
Never	7.97	3.24	0.42	0.46	0.50	0.49	0.46
Former	8.36	4.20	0.41	0.47	0.50	0.50	0.45
Current	9.99	3.53	0.43	0.49	0.52	0.54	0.49
Unknown	8.32	3.25	0.44	0.47	0.47	0.51	0.47
Living Status							
Deceased	8.15	4.24	0.43	0.47	0.52	0.51	0.48
Alive	8.24	3.37	0.42	0.46	0.50	0.50	0.46
Gender							
Male	8.22	4.53	0.42	0.45	0.50	0.50	0.46
Female	8.24	2.67	0.42	0.47	0.50	0.50	0.47

(Table Cont'd)

Study Characteristics	Average EP_UNINSUR	Average EP OCC_MBS	Average EP OCC_SER	Average EP OCC_SAL	Average EP OCC_NRC	Average EP OCC_PTM	Average Prim_RUCA
Race_Ethnicity							
White or Caucasian	9.10	35.05	16.74	21.49	10.80	12.60	1.56
Black or African American	10.88	26.76	20.33	20.44	9.40	13.78	1.51
Asian	8.73	40.18	15.33	22.46	9.25	12.57	1.19
Other	7.95	31.98	16.07	19.65	7.80	10.15	1.66
Unknown	10.63	31.99	19.39	20.78	10.64	12.29	2.28
Hispanic	12.92	29.16	19.72	21.88	11.33	13.88	1.31
Patient Type							
ED	10.70	26.62	19.15	20.58	9.55	13.67	1.55
Admit	10.12	28.50	19.65	20.55	10.13	13.25	1.72
Outpatient /UC	10.26	32.62	18.45	21.20	10.06	13.08	1.46
Smoking Status							
Never	10.24	29.86	19.02	20.92	9.89	13.36	1.47
Former	9.98	30.66	18.95	20.82	9.97	12.85	1.60
Current	10.13	31.08	18.74	20.77	8.78	13.05	1.43
Unknown	11.67	28.49	19.09	20.55	11.58	13.71	2.25
Living Status							
Deceased	10.24	29.54	18.97	20.61	10.15	13.42	1.52
Alive	10.31	30.05	18.99	20.88	9.95	13.24	1.57
Gender							
Male	10.19	30.32	19.01	20.84	10.26	13.07	1.68
Female	10.39	29.76	18.97	20.87	9.76	13.41	1.48

Appendix V. Summary of COVID Negative Patients Based on Demographics

Study Characteristics	Negative %	Alive %	Dead %	Male %	Female %	Average CPMI	Average Age	Average BMI
Race_Ethnicity								
White or Caucasian	57.89	56.44	1.45	25.15	32.74	45.40	23.22	58.38
Black or AfricanAmerican	33.45	32.59	0.86	14.44	19.02	52.99	25.79	52.96
Asian	0.59	0.59	0.00	0.25	0.34	47.77	16.49	52.85
Other	1.04	1.03	0.01	0.49	0.55	66.62	12.67	54.78
Unknown	5.37	5.32	0.05	2.33	3.03	75.30	5.12	52.79
Hispanic	1.66	1.66	0.01	0.85	0.81	60.71	19.91	46.60
Patient Type								
ED	14.29	14.15	0.15	5.99	8.31	86.63	29.80	45.85
Admit	27.75	25.81	1.95	14.15	13.61	49.87	29.06	62.96
Outpatient /UC	57.96	57.67	0.28	23.38	34.58	41.09	18.26	55.17
Smoking Status								
Never	47.66	46.70	0.96	16.92	30.74	47.15	25.15	54.94
Former	22.87	21.96	0.90	12.29	10.58	33.45	26.89	63.87
Current	18.89	18.39	0.50	9.67	9.22	58.61	24.63	51.54
Unknown	10.58	10.57	0.01	4.64	5.94	83.57	1.11	51.71
Age_Category								
11 to 20	2.16	2.14	0.01	0.82	1.34	77.15	18.17	19.69
21 to 30	10.19	10.16	0.03	4.17	6.02	75.21	21.04	26.25
31 to 40	12.53	12.43	0.11	4.73	7.81	67.29	22.79	36.12
41 to 50	13.99	13.83	0.16	5.31	8.68	64.14	24.05	46.13
51 to 60	17.64	17.35	0.29	7.96	9.68	54.42	23.56	56.32
61 to 70	20.13	19.56	0.57	9.91	10.22	43.81	23.48	65.87
71 to 80	14.13	13.50	0.63	6.99	7.14	26.56	24.20	75.59
81 to 90	7.31	6.85	0.45	3.03	4.28	14.01	20.51	85.38
91 to 100	1.87	1.76	0.12	0.60	1.28	5.27	18.25	93.99
101 to 110	0.04	0.03	0.01	0.01	0.03	0.90	11.43	101.56
Institutionalized								
No	91.58	89.57	2.01	39.30	52.28	51.53	23.48	54.45
Yes	8.42	8.06	0.36	4.21	4.21	33.80	16.67	72.86

(Table Cont'd)

Study Characteristics	Negative %	Alive %	Dead %	Male %	Female %	Average CPMI	Average Age	Average BMI
Institution Group								
NonInstitutionalized	91.58	89.57	2.01	39.30	52.28	51.53	23.48	54.45
Nursing Home	7.63	7.29	0.34	3.70	3.94	31.66	15.87	74.34
Rehab/Mental Health	0.16	0.15	0.01	0.05	0.11	23.81	15.32	72.77
Prison/Jail	0.63	0.62	0.01	0.47	0.16	62.44	26.73	54.96
Gender								
Male	43.51	42.21	1.31	43.51	0.00	49.46	22.33	57.08
Female	56.49	55.42	1.07	0.00	56.49	50.48	23.35	55.17
Living Status								
Deceased	2.38	0.00	2.38	1.31	1.07	32.17	27.54	69.58
Alive	97.62	97.62	0.00	42.21	55.42	50.47	22.79	55.67

Appendix W. Summary of Negative Patients Based on Underlying Conditions

Study Characteristics	% HTN	% Diabetes	% CVD	% CKD	% CEVD	% Obesity	%Respiratory	% Other Problems
Race_Ethnicity								
White or Caucasian	28.55	13.44	8.33	6.10	3.33	6.87	9.25	8.80
Black or African American	18.30	8.68	4.75	4.17	1.87	4.19	4.75	5.78
Asian	0.25	0.18	0.05	0.07	0.03	0.04	0.05	0.09
Other	0.41	0.15	0.12	0.05	0.05	0.09	0.09	0.07
Unknown	1.41	0.69	0.52	0.36	0.20	0.31	0.42	0.44
Hispanic	0.68	0.35	0.15	0.12	0.07	0.10	0.20	0.20
Patient Type								
ED	6.23	2.68	1.45	1.09	0.56	1.42	2.08	1.87
Admit	15.96	8.06	5.59	4.53	2.07	3.87	4.95	5.79
Outpatient /UC	27.40	12.75	6.88	5.26	2.91	6.31	7.73	7.73
Smoking Status								
Never	24.52	11.61	6.10	5.37	2.71	5.99	6.20	6.63
Former	13.55	6.47	4.47	3.14	1.51	3.11	4.62	4.60
Current	9.30	4.32	2.57	1.85	1.03	1.93	3.38	3.42
Unknown	2.21	1.09	0.78	0.51	0.30	0.58	0.56	0.74
Living Status								
Deceased	1.53	0.72	0.65	0.48	0.18	0.29	0.49	0.55
Alive	48.05	22.77	13.27	10.39	5.37	11.31	14.27	14.84
Gender								
Male	21.88	10.38	6.52	4.80	2.57	4.70	6.07	6.48
Female	27.71	13.10	7.40	6.07	2.97	6.91	8.69	8.91
Age Category								
11 to 20	0.50	0.28	0.18	0.15	0.09	0.17	0.26	0.24
21 to 30	2.84	1.50	0.85	0.75	0.31	1.01	1.28	1.09
31 to 40	4.28	1.94	0.98	0.81	0.41	1.45	1.48	1.65
41 to 50	6.35	2.77	1.54	1.23	0.62	1.74	1.74	2.00
51 to 60	8.96	4.32	2.35	1.82	1.04	2.19	2.59	3.02
61 to 70	11.75	5.73	3.15	2.25	1.30	2.41	3.50	3.57
71 to 80	8.87	4.69	2.96	2.35	1.04	1.81	2.44	2.30
81 to 90	4.81	1.87	1.53	1.25	0.59	0.69	1.19	1.19
91 to 100	1.18	0.38	0.36	0.25	0.14	0.13	0.28	0.32
101 to 110	0.03	0.01	0.01	0.01	0.01	0.01	0.00	0.01
Institutionalized								
No	44.55	20.99	12.43	9.71	4.79	10.81	13.55	13.90
Yes	5.04	2.49	1.49	1.17	0.75	0.79	1.21	1.49

(Table Cont'd)

Study Characteristics	% HTN	% Diabetes	% CVD	% CKD	% CEVD	% Obesity	%Respiratory	% Other Problems
Institution Group								
Non_Institutionalized	44.55	20.99	12.43	9.71	4.79	10.81	13.55	13.90
Nursing Home	4.64	2.31	1.38	1.08	0.73	0.74	1.11	1.33
Rehab/Mental Health	0.09	0.06	0.01	0.02	0.01	0.02	0.03	0.04
Prison/Jail	0.31	0.12	0.10	0.07	0.01	0.04	0.07	0.13

Appendix X. Summary of COVID Negative Patients Based on SVI and ACS Data

Study Characteristics	Average EP_POV	Average EP_UNEMP	Average EP_PCI	Average EP_NOHSDP	Average EP_AGE65	Average EP_AGE17
Race_Ethnicity						
White or Caucasian	15.14	5.50	29857.34	12.57	14.60	24.06
Black or African American	25.53	9.00	23022.96	17.05	13.59	24.82
Asian	13.40	6.14	33603.43	9.57	15.53	22.62
Other	16.20	5.83	28084.64	14.87	14.86	24.43
Unknown	20.25	6.85	26727.51	15.10	14.74	23.99
Hispanic	19.44	6.79	27046.50	13.54	13.54	23.94
Patient Type						
ED	21.89	7.75	25420.57	15.75	13.68	24.89
Admit	20.61	7.43	26141.99	15.67	14.07	24.24
Outpatient /UC	17.45	6.22	28421.49	13.16	14.49	24.19
Smoking Status						
Never	18.51	6.63	28006.27	13.58	14.18	24.28
Former	18.00	6.63	27698.24	14.20	14.27	24.26
Current	20.67	7.43	26004.09	15.47	13.82	24.55
Unknown	20.03	6.55	26138.84	14.95	15.35	24.04
Living Status						
Deceased	20.21	6.94	26798.80	14.77	14.38	23.97
Alive	18.93	6.77	27373.64	14.21	14.26	24.31
Gender						
Male	18.96	6.81	27208.90	14.34	14.29	24.19
Female	18.96	6.74	27476.34	14.14	14.23	24.39

(Table Cont'd)

Study Characteristics	Average EP_DISABL	Average EP_SNGPNT	Average EP_MINRTY	Average EP_LIMENG	Average EP_MUNIT	Average EP_MOBILE	Average EP_CROWD
Race_Ethnicity							
White or Caucasian	24.06	13.48	9.82	28.77	0.98	5.74	16.78
Black or African American	24.82	15.98	14.00	61.77	1.12	8.71	10.09
Asian	22.62	13.66	9.06	36.77	1.49	12.31	8.82
Other	24.43	14.43	10.11	33.77	1.32	5.85	16.74
Unknown	23.99	14.68	10.91	37.94	1.30	7.06	14.42
Hispanic	23.94	13.59	11.46	40.91	1.56	6.96	13.72
Patient Type							
ED	24.89	14.31	12.74	44.06	0.92	6.47	14.30
Admit	24.24	15.15	11.81	45.38	1.10	7.08	13.79
Outpatient /UC	24.19	14.05	10.71	37.46	1.07	6.85	14.57
Smoking Status							
Never	24.28	14.04	11.23	40.89	1.06	7.14	13.67
Former	24.26	14.57	11.04	39.47	1.04	6.52	14.96
Current	24.55	15.07	12.19	44.28	1.09	7.37	14.36
Unknown	24.04	14.37	10.60	35.16	1.03	5.43	15.79
Living Status							
Deceased	23.97	15.35	11.45	45.19	1.29	9.07	12.53
Alive	24.31	14.37	11.30	40.49	1.05	6.81	14.36
Gender							
Male	24.19	14.43	11.31	40.85	1.02	6.74	14.50
Female	24.39	14.36	11.29	40.41	1.09	6.96	14.18

(Table Cont'd)

Study Characteristics	Average EP_NOVEH	Average EP_GROUPQ	Average RPL THEME1	Average RPL THEME2	Average RPL THEME3	Average RPL THEME4	Average RPL THEMES
Race_Ethnicity							
White or Caucasian	5.71	1.85	0.34	0.41	0.37	0.46	0.35
Black or African American	11.52	3.87	0.57	0.56	0.56	0.60	0.60
Asian	7.12	1.63	0.29	0.39	0.50	0.41	0.34
Other	5.95	1.78	0.40	0.48	0.45	0.52	0.44
Unknown	7.44	2.58	0.46	0.48	0.45	0.54	0.48
Hispanic	7.21	3.26	0.42	0.43	0.49	0.52	0.45
Patient Type							
ED	9.24	3.00	0.49	0.50	0.45	0.54	0.50
Admit	8.66	3.20	0.47	0.49	0.47	0.55	0.49
Outpatient /UC	7.01	2.19	0.38	0.45	0.42	0.48	0.41
Smoking Status							
Never	7.57	2.50	0.41	0.45	0.44	0.50	0.43
Former	7.52	2.47	0.41	0.47	0.44	0.50	0.43
Current	8.91	2.88	0.47	0.50	0.46	0.54	0.49
Unknown	7.32	2.70	0.45	0.48	0.40	0.53	0.45
Living Status							
Deceased	9.34	2.97	0.44	0.49	0.51	0.54	0.48
Alive	7.75	2.58	0.42	0.47	0.44	0.51	0.44
Gender							
Male	7.83	2.93	0.43	0.47	0.44	0.51	0.45
Female	7.75	2.33	0.42	0.47	0.45	0.51	0.44

(Table Cont'd)

Study Characteristics	Average EP_UNINSUR	Average EP OCC_MBS	Average EP OCC_SER	Average EP OCC_SAL	Average EP OCC_NRC	Average EP OCC_PTM	Average Prim_RUCA
Race_Ethnicity							
White or Caucasian	9.63	35.18	16.57	22.72	12.03	13.47	1.85
Black or African American	12.22	28.49	23.20	22.80	10.13	15.30	1.67
Asian	8.85	40.45	16.21	22.22	9.31	11.77	1.27
Other	9.67	33.03	17.83	22.72	12.62	13.75	2.31
Unknown	11.12	33.16	19.08	22.31	11.61	13.82	2.69
Hispanic	11.30	33.44	18.78	22.43	11.51	13.81	1.71
Patient Type							
ED	11.39	30.73	20.11	23.39	10.90	14.86	1.60
Admit	10.99	31.02	20.16	22.56	11.66	14.50	1.87
Outpatient /UC	10.22	34.19	18.12	22.63	11.32	13.73	1.87
Smoking Status							
Never	10.40	33.53	18.72	22.79	10.98	13.91	1.73
Former	10.37	32.98	18.70	22.61	11.73	13.97	1.78
Current	11.15	30.85	20.19	22.58	11.56	14.80	1.69
Unknown	11.02	32.77	18.52	22.89	11.86	13.98	2.65
Living Status							
Deceased	0.48	10.99	31.91	20.34	21.96	11.52	14.23
Alive	0.44	10.59	32.84	18.94	22.74	11.35	14.10
Gender							
Male	0.45	10.60	32.80	18.96	22.72	11.38	14.09
Female	0.44	10.60	32.83	18.98	22.72	11.33	14.11

Appendix Y. Summary of COVID Positive Institutionalized Patients Based on Demographics

Study Characteristics	Positive %	Alive %	Dead %	Male %	Female %	Average CPMI	Average BMI	Average Age
Race_Ethnicity								
White or Caucasian	46.39	36.28	10.11	19.13	27.26	22.33	25.28	77.10
Black or African American	43.68	34.84	8.84	24.73	18.95	24.00	33.29	70.34
Asian	0.54	0.18	0.36	0.54	0.00	16.37	25.83	67.49
Other	1.26	1.08	0.18	0.36	0.90	15.51	40.95	78.69
Unknown	6.86	6.32	0.54	3.43	3.43	12.97	43.05	74.33
Hispanic	1.26	1.08	0.18	0.90	0.36	21.01	34.95	65.59
Patient Type								
ED	3.79	3.07	0.72	1.99	1.81	28.52	54.74	71.28
Admit	50.72	35.92	14.80	28.70	22.02	29.90	33.93	70.90
Outpatient /UC	45.49	40.79	4.69	18.41	27.08	13.26	24.25	77.20
Smoking Status								
Never	48.92	39.89	9.03	19.13	29.78	28.65	21.59	76.45
Former	31.23	22.74	8.48	17.33	13.90	26.61	25.48	73.47
Current	13.18	11.01	2.17	8.48	4.69	39.21	22.33	63.91
Unknown	6.68	6.14	0.54	4.15	2.53	42.31	12.31	75.15
Age Category								
21 to 30	0.36	0.36	0.00	0.18	0.18	28.61	90.15	28.24
31 to 40	1.26	1.26	0.00	0.72	0.54	29.48	88.56	34.23
41 to 50	3.79	3.43	0.36	2.53	1.26	28.18	81.94	46.14
51 to 60	10.47	8.48	1.99	6.50	3.97	25.54	62.61	57.07
61 to 70	24.73	19.13	5.60	15.16	9.57	23.66	44.50	66.54
71 to 80	25.27	20.04	5.23	13.18	12.09	23.24	20.61	75.75
81 to 90	25.63	20.40	5.23	9.57	16.06	20.26	10.00	85.70
91 to 100	7.94	6.14	1.81	1.08	6.86	14.03	5.52	93.18
101 to 110	0.54	0.54	0.00	0.18	0.36	6.16	0.00	102.81
Institution Group								
Nursing Home	89.17	70.40	18.77	40.43	48.74	25.93	21.35	75.55
Rehab/Mental Health	1.44	1.08	0.36	1.44	0.00	56.22	27.16	71.97
Prison/Jail	9.39	8.30	1.08	7.22	2.17	67.98	30.35	57.19
Gender								
Male	49.10	36.64	12.45	49.10	0.00	38.04	21.92	70.34
Female	50.90	43.14	7.76	0.00	50.90	22.87	22.63	77.10
Living Status								
Deceased	20.22	0.00	20.22	12.45	7.76	25.32	27.41	75.28
Alive	79.78	79.78	0.00	36.64	43.14	31.58	20.98	73.40

Appendix Z. Summary of Institutionalized Positive Patients Based on Underlying Conditions

Study Characteristics	% HTN	% Diabetes	% CVD	% CKD	% CEVD	% Obesity	%Respiratory	% PVD	% Other Problems
Race_Ethnicity									
White or Caucasian	40.61	19.49	17.69	14.26	11.37	9.57	12.09	3.25	6.32
Black or African American	38.63	22.56	16.43	13.72	13.72	8.66	7.04	5.05	8.66
Asian	0.54	0.36	0.18	0.18	0.54	0.00	0.18	0.00	0.18
Other	0.90	0.18	0.36	0.18	0.18	0.00	0.00	0.00	0.00
Unknown	4.69	1.99	1.44	1.44	0.90	0.54	0.72	0.36	0.54
Hispanic	1.08	0.72	0.36	0.72	0.00	0.36	0.18	0.18	0.00
Patient Type									
ED	2.53	1.62	0.90	0.72	1.26	0.54	0.54	0.36	0.00
Admit	42.96	25.63	21.12	16.97	15.52	12.82	12.45	5.96	9.93
Outpatient /UC	40.97	18.05	14.44	12.82	9.93	5.78	7.22	2.53	5.78
Smoking Status									
Never	43.32	22.56	18.23	15.34	11.37	9.75	7.22	3.25	5.78
Former	26.71	14.62	12.27	10.47	9.57	7.04	6.86	3.61	5.78
Current	12.64	6.68	4.69	3.79	4.15	1.99	5.78	1.99	3.97
Unknown	3.79	1.44	1.26	0.90	1.62	0.36	0.36	0.00	0.18
Living Status									
Deceased	17.51	10.65	9.93	9.03	6.50	4.15	5.60	2.53	3.25
Alive	68.95	34.66	26.53	21.48	20.22	14.98	14.62	6.32	12.45
Age Category									
21 to 30	0.00	0.00	0.00	0.00	0.18	0.18	0.00	0.00	0.00
31 to 40	0.54	0.36	0.00	0.00	0.00	0.36	0.18	0.00	0.00
41 to 50	2.71	0.90	0.72	0.36	0.18	1.99	0.54	0.18	0.36
51 to 60	9.75	4.87	3.07	1.99	3.79	3.43	1.99	0.36	3.07
61 to 70	21.66	13.54	8.48	8.30	8.66	5.78	4.69	2.35	5.60
71 to 80	22.92	13.18	11.19	8.84	7.94	5.23	7.04	3.61	3.97
81 to 90	21.84	9.75	9.75	8.30	4.87	1.81	4.33	1.62	2.35
91 to 100	6.50	2.71	3.07	2.53	0.90	0.36	1.44	0.72	0.18
101 to 110	0.54	0.00	0.18	0.18	0.18	0.00	0.00	0.00	0.18
Institution Group									
Nursing Home	79.42	42.42	34.48	28.70	25.45	15.88	18.41	8.66	13.72
Rehab/Mental Health	0.90	0.54	0.36	0.36	0.18	0.00	0.36	0.00	0.00
Prison/Jail	6.14	2.35	1.62	1.44	1.08	3.25	1.44	0.18	1.99

Appendix AA. Summary of COVID Positive Non Institutionalized Patients Based on Demographics

Study Characteristics	Positive %	Alive %	Dead %	Male %	Female %	Average CPMI	Average BMI	Average Age
Race_Ethnicity								
White or Caucasian	29.63	27.31	2.31	14.02	15.61	53.40	22.54	55.97
Black or African American	59.48	54.90	4.58	21.73	37.75	63.74	29.44	50.26
Asian	0.59	0.59	0.00	0.23	0.36	70.53	17.67	46.16
Other	1.27	1.13	0.14	0.64	0.64	76.48	17.36	48.04
Unknown	4.81	4.63	0.18	2.27	2.54	79.07	7.34	45.67
Hispanic	4.22	4.22	0.00	2.04	2.18	77.06	24.17	40.75
Patient Type								
ED	25.64	25.14	0.50	9.66	15.97	89.78	31.87	43.69
Admit	28.13	21.73	6.40	13.29	14.84	56.49	32.45	62.11
Outpatient /UC	46.23	45.92	0.32	17.97	28.27	50.34	18.57	48.89
Smoking Status								
Never	66.47	62.07	4.40	22.96	43.51	62.83	28.26	49.46
Former	17.88	15.29	2.59	9.62	8.26	46.66	29.12	61.99
Current	6.26	5.94	0.32	3.86	2.40	66.29	26.12	47.62
Unknown	9.39	9.30	0.09	4.49	4.90	84.38	2.76	46.18
Age Category								
11 to 20	3.54	3.49	0.05	1.45	2.09	85.36	19.89	19.54
21 to 30	13.97	13.88	0.09	5.31	8.71	75.64	23.69	26.38
31 to 40	14.34	14.16	0.18	5.49	9.03	76.33	26.62	35.96
41 to 50	16.70	16.24	0.45	7.08	10.25	67.35	27.57	46.36
51 to 60	19.92	18.47	1.45	9.35	12.21	67.20	26.84	56.18
61 to 70	16.74	15.20	1.54	11.71	8.85	53.20	26.42	65.48
71 to 80	8.98	7.12	1.86	7.80	4.49	36.89	25.9	75.29
81 to 90	4.40	3.09	1.32	4.22	2.59	15.91	24.64	85.38
91 to 100	1.41	0.95	0.45	0.82	0.86	6.64	19.34	94.33
Gender								
Male	40.93	36.80	4.13	40.93	0.00	59.68	24.18	52.36
Female	59.07	55.81	3.27	0.00	59.07	63.92	27.07	50.53
Living Status								
Deceased	7.40	0.00	7.40	4.13	3.27	43.78	32.17	68.98
Alive	92.60	92.60	0.00	36.80	55.81	63.65	25.38	49.86

Appendix AB. Summary of Non-Institutionalized Positive Patients Based on Underlying Conditions

Study Characteristics	% HTN	% Diabetes	% CVD	% CKD	% CEVD	% Obesity	%Respiratory	% Other Problems
Race_Ethnicity								
White or Caucasian	13.66	6.58	4.13	2.63	1.77	6.03	3.18	3.27
Black or African American	33.62	20.78	8.44	7.76	4.08	18.87	6.58	9.53
Asian	0.05	0.05	0.00	0.00	0.00	0.05	0.05	0.00
Other	0.27	0.18	0.05	0.09	0.05	0.18	0.05	0.18
Unknown	0.50	0.09	0.09	0.09	0.00	0.41	0.18	0.23
Hispanic	0.68	0.64	0.14	0.14	0.09	0.91	0.14	0.14
Patient Type								
ED	10.34	5.44	2.04	1.23	0.91	4.63	2.09	1.77
Admit	21.10	14.11	8.44	7.71	4.13	11.03	5.22	7.76
Outpatient /UC	17.33	8.76	2.36	1.77	0.95	10.80	2.86	3.81
Smoking Status								
Never	26.13	15.83	6.85	5.90	3.27	15.29	5.40	6.94
Former	8.98	4.63	2.86	1.72	1.00	4.13	1.77	2.99
Current	8.80	4.85	2.04	1.81	1.13	4.45	1.81	2.09
Unknown	4.85	2.99	1.09	1.27	0.59	2.59	1.18	1.32
Living Status								
Deceased	5.81	4.36	2.59	2.68	1.41	3.04	1.50	2.31
Alive	42.97	23.96	10.25	8.03	4.58	23.41	8.67	11.03
Gender								
Male	20.46	12.43	6.99	6.08	3.18	8.76	3.81	5.81
Female	28.31	15.88	5.85	4.63	2.81	17.70	6.35	7.53
Age Category								
11 to 20	0.18	0.14	0.00	0.00	0.05	0.41	0.18	0.09
21 to 30	1.09	0.73	0.05	0.09	0.05	1.91	0.95	0.36
31 to 40	3.13	1.91	0.23	0.45	0.09	3.77	0.86	1.04
41 to 50	8.48	4.36	0.95	1.09	0.50	5.58	1.45	1.68
51 to 60	12.02	6.94	2.31	2.09	1.09	6.90	2.22	4.08
61 to 70	11.98	7.80	3.99	3.09	1.95	5.17	2.18	2.77
71 to 80	7.35	4.17	2.95	2.04	1.18	2.00	1.27	2.13
81 to 90	3.40	1.86	1.95	1.32	0.95	0.68	0.86	0.82
91 to 100	1.13	0.41	0.41	0.54	0.14	0.05	0.18	0.36

Appendix AC. Summary of COVID Positive Non-Institutionalized Patients Based on SVI and ACS Data

Study Characteristics	Average EP_POV	Average EP_UNEMP	Average EP_PCI	Average EP_NOHSDP	Average EP_AGE65	Average EP_AGE17
Race_Ethnicity						
White or Caucasian	13.89	5.13	29478.69	11.34	13.50	23.39
Black or African American	22.00	8.07	20716.56	15.37	11.85	22.52
Asian	13.72	4.84	31079.08	11.07	13.30	24.41
Other	17.10	6.35	22138.82	11.09	11.36	19.10
Unknown	19.42	6.36	24655.11	14.30	13.73	23.20
Hispanic	18.04	6.37	25241.70	13.19	11.99	23.33
Patient Type						
ED	21.34	7.52	21708.80	15.19	11.96	22.46
Admit	20.21	7.48	21982.33	14.51	12.05	22.21
Outpatient /UC	17.39	6.43	26005.21	12.94	12.94	23.37
Smoking Status						
Never	19.08	7.09	23895.22	13.85	12.29	22.97
Former	18.90	7.03	24048.04	13.88	12.58	22.48
Current	21.53	7.65	22722.17	14.05	11.88	22.34
Unknown	19.01	5.94	23076.08	14.75	13.42	22.64
Living Status						
Deceased	19.91	7.22	21411.54	14.43	12.52	21.45
Alive	19.14	6.99	23960.68	13.92	12.41	22.92
Gender						
Male	18.60	6.84	24664.17	13.70	12.51	22.76
Female	19.61	7.12	23154.19	14.13	12.39	22.85

(Table Cont'd)

Study Characteristics	Average EP_DISABL	Average EP_SNGPNT	Average EP_MINRTY	Average EP_LIMENG	Average EP_MUNIT	Average EP_MOBILE	Average EP_CROWD
Race_Ethnicity							
White or Caucasian	12.48	9.67	28.54	0.86	5.56	14.94	2.12
Black or African American	14.12	12.67	56.94	1.07	6.16	9.55	2.76
Asian	12.73	11.05	37.67	1.42	5.19	16.68	2.62
Other	11.21	8.82	34.69	0.45	6.84	8.42	2.41
Unknown	13.63	10.88	42.65	1.15	5.91	13.98	2.44
Hispanic	12.05	11.09	42.19	2.77	6.55	14.05	2.57
Patient Type							
ED	13.01	12.47	47.07	0.99	5.17	11.72	2.46
Admit	14.14	11.72	51.04	1.04	5.99	10.38	2.84
Outpatient /UC	13.34	10.98	44.11	1.15	6.44	12.23	2.40
Smoking Status							
Never	13.30	11.71	48.09	1.10	6.19	11.12	2.54
Former	14.01	11.38	45.13	0.90	5.62	11.86	2.50
Current	13.79	12.08	49.47	0.97	6.77	8.94	2.76
Unknown	13.52	10.58	39.28	1.30	4.76	16.02	2.44
Living Status							
Deceased	13.77	11.48	48.88	0.97	4.94	11.55	2.65
Alive	13.45	11.58	46.65	1.09	6.07	11.58	2.53
Gender							
Male	13.37	11.25	44.65	1.11	5.88	11.98	2.41
Female	13.55	11.80	48.32	1.06	6.07	11.30	2.63

(Table Cont'd)

Study Characteristics	Average EP_NOVEH	Average EP_GROUPQ	Average RPL THEME1	Average RPL THEME2	Average RPL THEME3	Average RPL THEME4	Average RPL THEMES
Race_Ethnicity							
White or Caucasian	4.91	1.49	0.31	0.40	0.37	0.41	0.33
Black or African American	9.11	2.81	0.51	0.51	0.53	0.52	0.54
Asian	5.85	3.84	0.26	0.39	0.49	0.49	0.35
Other	5.70	2.39	0.39	0.36	0.37	0.46	0.39
Unknown	7.62	2.38	0.43	0.47	0.45	0.47	0.45
Hispanic	6.12	2.06	0.41	0.37	0.54	0.48	0.42
Patient Type							
ED	8.77	3.18	0.48	0.48	0.47	0.52	0.51
Admit	8.06	1.88	0.47	0.49	0.49	0.49	0.49
Outpatient /UC	6.69	2.21	0.39	0.45	0.47	0.46	0.42
Smoking Status							
Never	7.53	2.55	0.44	0.46	0.48	0.48	0.46
Former	7.64	1.69	0.43	0.48	0.46	0.48	0.46
Current	9.12	2.63	0.47	0.49	0.50	0.53	0.51
Unknown	7.08	2.18	0.44	0.47	0.44	0.50	0.46
Living Status							
Deceased	7.93	2.32	0.46	0.48	0.48	0.48	0.49
Alive	7.58	2.37	0.44	0.47	0.48	0.48	0.46
Gender							
Male	7.33	2.12	0.42	0.45	0.48	0.47	0.45
Female	7.80	2.54	0.45	0.48	0.48	0.49	0.48

(Table Cont'd)

Study Characteristics	Average EP_UNINSUR	Average EP OCC_MBS	Average EP OCC_SER	Average EP OCC_SAL	Average EP OCC_NRC	Average EP OCC_PTM	Ave Prim_RUCA
Race_Ethnicity							
White or Caucasian	8.54	34.18	15.71	21.71	10.85	12.94	1.59
Black or African American	10.80	24.81	20.31	20.34	9.40	14.21	1.44
Asian	8.72	38.34	14.63	23.55	10.00	13.25	1.23
Other	7.61	29.00	14.65	20.33	7.57	10.53	1.50
Unknown	10.80	29.67	18.84	21.37	11.17	13.25	2.25
Hispanic	13.01	29.21	19.53	21.65	11.42	13.86	1.30
Patient Type							
ED	10.72	26.34	19.11	20.53	9.48	13.71	1.53
Admit	10.19	25.90	19.48	20.07	9.97	13.42	1.48
Outpatient /UC	9.86	30.50	18.07	21.54	10.27	13.91	1.53
Smoking Status							
Never	10.12	28.42	18.95	20.93	9.82	13.74	1.45
Former	9.70	27.87	18.26	20.94	9.88	13.38	1.44
Current	9.92	27.35	18.66	20.55	9.03	13.60	1.43
Unknown	11.59	27.23	18.21	20.51	11.98	14.33	2.18
Living Status							
Deceased	9.88	25.05	19.00	19.58	10.12	13.95	1.46
Alive	10.19	28.39	18.72	20.97	9.97	13.70	1.52
Gender							
Male	10.00	28.68	18.42	20.93	10.21	13.54	1.51
Female	10.29	27.76	18.96	20.83	9.83	13.85	1.52

Appendix AD. Hot Spot Analysis FMOLHS Positive Institutionalized Patients Counts By Census Tracts across LA

Census Tract ID	Parish	Positive Count	Confidence Result
11.02	East Baton Rouge	15	Hot Spot 99%
17	East Baton Rouge	6	Hot Spot 95%
18	East Baton Rouge	86	Hot Spot 95%
202	West Baton Rouge	2	Hot Spot 95%
24	East Baton Rouge	5	Hot Spot 95%
26.02	East Baton Rouge	16	Hot Spot 95%
302.04	Ascension	4	Hot Spot 95%
303	Ascension	7	Hot Spot 95%
304.02	Ascension	1	Hot Spot 95%
310	Ascension	28	Hot Spot 95%
37.02	East Baton Rouge	8	Hot Spot 95%
37.03	East Baton Rouge	1	Hot Spot 95%
38.02	East Baton Rouge	5	Hot Spot 95%
38.04	East Baton Rouge	2	Hot Spot 95%
38.05	East Baton Rouge	7	Hot Spot 95%
39.09	East Baton Rouge	2	Hot Spot 95%
40.16	East Baton Rouge	16	Hot Spot 95%
406	Livingston	21	Hot Spot 95%
407	Livingston	15	Hot Spot 95%
408.04	Livingston	1	Hot Spot 95%
45.03	East Baton Rouge	8	Hot Spot 95%
45.04	East Baton Rouge	22	Hot Spot 95%
45.05	East Baton Rouge	1	Hot Spot 95%
45.09	East Baton Rouge	13	Hot Spot 95%
45.10	East Baton Rouge	20	Hot Spot 90%
48	East Baton Rouge	6	Hot Spot 90%
53	East Baton Rouge	1	Hot Spot 90%
9529	Iberville	1	Hot Spot 90%
9531.01	Iberville	48	Hot Spot 90%
9531.02	Iberville	22	Hot Spot 90%
9532	Iberville	2	Hot Spot 90%

Appendix AE. Hot Spot Analysis FMOLHS Positive Non-Institutionalized Patients Counts By Census Tracts across LA

Census Tract ID	Parish	Positive Count	Confidence Result
1	East Baton Rouge	4	Hot Spot 99%
10	East Baton Rouge	7	Hot Spot 99%
11.02	East Baton Rouge	2	Hot Spot 99%
11.03	East Baton Rouge	4	Hot Spot 99%
11.04	East Baton Rouge	5	Hot Spot 99%
16	East Baton Rouge	3	Hot Spot 99%
17	East Baton Rouge	2	Hot Spot 99%
18	East Baton Rouge	4	Hot Spot 99%
19	East Baton Rouge	4	Hot Spot 99%
2	East Baton Rouge	12	Hot Spot 99%
20	East Baton Rouge	2	Hot Spot 99%
201	West Baton Rouge	7	Hot Spot 99%
202	West Baton Rouge	2	Hot Spot 99%
203	West Baton Rouge	4	Hot Spot 99%
204.01	West Baton Rouge	16	Hot Spot 99%
204.02	West Baton Rouge	14	Hot Spot 99%
205	Lafourche	1	Hot Spot 99%
210	St. Martin	1	Hot Spot 99%
22	East Baton Rouge	4	Hot Spot 99%
23	East Baton Rouge	1	Hot Spot 99%
24	East Baton Rouge	5	Hot Spot 99%
25	East Baton Rouge	3	Hot Spot 99%
26.01	East Baton Rouge	3	Hot Spot 99%
26.02	East Baton Rouge	7	Hot Spot 99%
27	East Baton Rouge	3	Hot Spot 99%
28.02	East Baton Rouge	1	Hot Spot 99%
3	East Baton Rouge	6	Hot Spot 99%
3	Terrebonne	1	Hot Spot 99%
30	East Baton Rouge	3	Hot Spot 99%
301	Iberia	4	Hot Spot 99%
301.01	Ascension	9	Hot Spot 99%
301.02	Ascension	14	Hot Spot 99%
301.03	Ascension	4	Hot Spot 99%
302.03	Ascension	5	Hot Spot 99%
302.04	Ascension	21	Hot Spot 99%
302.05	Ascension	24	Hot Spot 99%
302.06	Ascension	22	Hot Spot 99%
303	Ascension	62	Hot Spot 99%

(Table Cont'd)

Census Tract ID	Parish	Positive Count	Confidence Result
304.01	Ascension	22	Hot Spot 99%
304.02	Ascension	57	Hot Spot 99%
305	Ascension	8	Hot Spot 99%
306	Ascension	25	Hot Spot 99%
309	Ascension	17	Hot Spot 99%
31.01	East Baton Rouge	8	Hot Spot 99%
310	Ascension	40	Hot Spot 99%
32.01	East Baton Rouge	7	Hot Spot 99%
32.02	East Baton Rouge	5	Hot Spot 99%
33	East Baton Rouge	16	Hot Spot 99%
34	East Baton Rouge	18	Hot Spot 99%
35.01	East Baton Rouge	5	Hot Spot 99%
35.04	East Baton Rouge	16	Hot Spot 99%
35.05	East Baton Rouge	13	Hot Spot 99%
35.06	East Baton Rouge	21	Hot Spot 99%
35.07	East Baton Rouge	14	Hot Spot 99%
36.01	East Baton Rouge	4	Hot Spot 99%
36.03	East Baton Rouge	4	Hot Spot 99%
36.04	East Baton Rouge	10	Hot Spot 99%
37.01	East Baton Rouge	3	Hot Spot 99%
37.02	East Baton Rouge	7	Hot Spot 99%
37.03	East Baton Rouge	7	Hot Spot 99%
38.01	East Baton Rouge	12	Hot Spot 99%
38.02	East Baton Rouge	3	Hot Spot 99%
38.04	East Baton Rouge	3	Hot Spot 99%
38.05	East Baton Rouge	9	Hot Spot 99%
39.04	East Baton Rouge	7	Hot Spot 99%
39.06	East Baton Rouge	3	Hot Spot 99%
39.07	East Baton Rouge	5	Hot Spot 99%
39.08	East Baton Rouge	2	Hot Spot 99%
39.09	East Baton Rouge	4	Hot Spot 99%
39.1	East Baton Rouge	4	Hot Spot 99%
4	East Baton Rouge	11	Hot Spot 99%
40.05	East Baton Rouge	10	Hot Spot 99%
40.06	East Baton Rouge	12	Hot Spot 99%
40.09	East Baton Rouge	8	Hot Spot 99%
40.1	East Baton Rouge	2	Hot Spot 99%
40.11	East Baton Rouge	9	Hot Spot 99%
40.13	East Baton Rouge	1	Hot Spot 99%
40.14	East Baton Rouge	7	Hot Spot 99%
40.15	East Baton Rouge	14	Hot Spot 99%
40.16	East Baton Rouge	9	Hot Spot 99%

(Table Cont'd)

Census Tract ID	Parish	Positive Count	Confidence Result
401	Livingston	1	Hot Spot 99%
401	St. James	10	Hot Spot 99%
402	St. James	4	Hot Spot 99%
402	St. Mary	1	Hot Spot 99%
402.01	Livingston	2	Hot Spot 99%
402.02	Livingston	2	Hot Spot 99%
403	St. James	3	Hot Spot 99%
403.01	Livingston	8	Hot Spot 99%
403.03	Livingston	7	Hot Spot 99%
403.04	Livingston	5	Hot Spot 99%
404	St. James	14	Hot Spot 99%
404.01	Livingston	5	Hot Spot 99%
404.02	Livingston	7	Hot Spot 99%
405	Livingston	7	Hot Spot 99%
405	St. James	17	Hot Spot 99%
406	Livingston	7	Hot Spot 99%
406	St. James	6	Hot Spot 99%
407	Livingston	4	Hot Spot 99%
407	St. James	2	Hot Spot 99%
408.02	Livingston	4	Hot Spot 99%
408.04	Livingston	15	Hot Spot 99%
408.05	Livingston	3	Hot Spot 99%
408.06	Livingston	5	Hot Spot 99%
409.02	Livingston	4	Hot Spot 99%
414	St. Mary	1	Hot Spot 99%
42.01	East Baton Rouge	12	Hot Spot 99%
42.04	East Baton Rouge	2	Hot Spot 99%
42.05	East Baton Rouge	11	Hot Spot 99%
43.01	East Baton Rouge	5	Hot Spot 99%
43.02	East Baton Rouge	2	Hot Spot 99%
44.01	East Baton Rouge	7	Hot Spot 99%
44.02	East Baton Rouge	4	Hot Spot 99%
44.03	East Baton Rouge	3	Hot Spot 99%
45.03	East Baton Rouge	8	Hot Spot 99%
45.04	East Baton Rouge	8	Hot Spot 99%
45.05	East Baton Rouge	9	Hot Spot 99%
45.07	East Baton Rouge	6	Hot Spot 99%
45.08	East Baton Rouge	6	Hot Spot 99%
45.09	East Baton Rouge	6	Hot Spot 99%
45.1	East Baton Rouge	9	Hot Spot 99%
46.02	East Baton Rouge	8	Hot Spot 99%
46.03	East Baton Rouge	4	Hot Spot 99%

(Table Cont'd)

Census Tract ID	Parish	Positive Count	Confidence Result
46.04	East Baton Rouge	13	Hot Spot 99%
47	East Baton Rouge	5	Hot Spot 99%
48	East Baton Rouge	1	Hot Spot 99%
49	East Baton Rouge	2	Hot Spot 99%
5	East Baton Rouge	11	Hot Spot 99%
50	East Baton Rouge	2	Hot Spot 99%
501	Assumption	29	Hot Spot 99%
502	Assumption	3	Hot Spot 99%
503	Assumption	38	Hot Spot 99%
504	Assumption	24	Hot Spot 99%
505	Assumption	11	Hot Spot 99%
506	Assumption	5	Hot Spot 99%
51	East Baton Rouge	3	Hot Spot 99%
52	East Baton Rouge	8	Hot Spot 99%
53	East Baton Rouge	4	Hot Spot 99%
6.01	East Baton Rouge	7	Hot Spot 99%
6.02	East Baton Rouge	13	Hot Spot 99%
7.01	East Baton Rouge	2	Hot Spot 99%
7.02	East Baton Rouge	10	Hot Spot 99%
702	St. John the Baptist	2	Hot Spot 99%
705	St. John the Baptist	1	Hot Spot 99%
707	St. John the Baptist	1	Hot Spot 99%
708	St. John the Baptist	1	Hot Spot 99%
709	St. John the Baptist	1	Hot Spot 99%
711	St. John the Baptist	1	Hot Spot 99%
9	East Baton Rouge	9	Hot Spot 99%
9512	St. Helena	1	Hot Spot 99%
9515.01	East Feliciana	3	Hot Spot 99%
9515.02	East Feliciana	9	Hot Spot 99%
9516	East Feliciana	3	Hot Spot 99%
9522	Pointe Coupee	2	Hot Spot 99%
9527	Iberville	12	Hot Spot 99%
9529	Iberville	22	Hot Spot 99%
9530	Iberville	3	Hot Spot 99%
9531.01	Iberville	10	Hot Spot 99%
9531.02	Iberville	7	Hot Spot 99%
9532	Iberville	15	Hot Spot 99%
9541.02	Tangipahoa	1	Hot Spot 99%
9544	Tangipahoa	1	Hot Spot 99%
2	West Carroll	7	Hot Spot 95%
201	St. Martin	4	Hot Spot 95%
3	West Carroll	1	Hot Spot 95%

(Table Cont'd)

Census Tract ID	Parish	Positive Count	Confidence Result
410	St. Mary	2	Hot Spot 95%
7	Terrebonne	1	Hot Spot 95%
8	Terrebonne	1	Hot Spot 95%
9502	Morehouse	5	Hot Spot 95%
9503	Morehouse	7	Hot Spot 95%
9504	Morehouse	4	Hot Spot 95%
9505	Morehouse	2	Hot Spot 95%
9506	Morehouse	4	Hot Spot 95%
9526	Iberville	3	Hot Spot 95%
9601	Lincoln	2	Hot Spot 95%
9601	Union	12	Hot Spot 95%
9602	Union	4	Hot Spot 95%
9603	Union	1	Hot Spot 95%
9604	Union	7	Hot Spot 95%
9606	Lincoln	1	Hot Spot 95%
9609	Lincoln	1	Hot Spot 95%
102.02	Ouachita	24	Hot Spot 90%
104	Ouachita	10	Hot Spot 90%
105.03	Ouachita	3	Hot Spot 90%
2	Caldwell	1	Hot Spot 90%
3	Caldwell	2	Hot Spot 90%
9502	Franklin	1	Hot Spot 90%
9507	Morehouse	1	Hot Spot 90%
9508	Morehouse	13	Hot Spot 90%
9546	Tangipahoa	1	Hot Spot 90%
9605	Union	8	Hot Spot 90%
9607	Lincoln	2	Hot Spot 90%
9701	Jackson	1	Hot Spot 90%
9701	Richland	2	Hot Spot 90%
9703	Jackson	3	Hot Spot 90%
9703	Richland	6	Hot Spot 90%
9705	Richland	11	Hot Spot 90%
1	Catahoula	1	Cold Spot 90%
2	Tensas	2	Cold Spot 90%
204.01	Grant	1	Cold Spot 90%
9507	Washington	1	Cold Spot 90%
120	Rapides	1	Cold Spot 95%
125	Rapides	1	Cold Spot 95%
218.04	Jefferson	1	Cold Spot 95%
25.03	Orleans	1	Cold Spot 95%
269	Jefferson	1	Cold Spot 95%
277.01	Jefferson	1	Cold Spot 95%

(Table Cont'd)

Census Tract ID	Parish	Positive Count	Confidence Result
306	Iberia	2	Cold Spot 95%
410.03	St. Tammany	1	Cold Spot 95%
9506	Washington	1	Cold Spot 95%
1	Lafayette	3	Cold Spot 99%
10.01	Lafayette	5	Cold Spot 99%
10.02	Lafayette	2	Cold Spot 99%
10.03	Lafayette	3	Cold Spot 99%
11	Lafayette	6	Cold Spot 99%
12	Lafayette	9	Cold Spot 99%
13	Lafayette	2	Cold Spot 99%
14.01	Lafayette	9	Cold Spot 99%
14.02	Lafayette	5	Cold Spot 99%
14.03	Lafayette	6	Cold Spot 99%
14.04	Lafayette	10	Cold Spot 99%
14.05	Lafayette	7	Cold Spot 99%
14.06	Lafayette	10	Cold Spot 99%
14.07	Lafayette	2	Cold Spot 99%
14.09	Lafayette	5	Cold Spot 99%
14.1	Lafayette	14	Cold Spot 99%
17	Lafayette	3	Cold Spot 99%
17.25	Orleans	1	Cold Spot 99%
18.01	Lafayette	1	Cold Spot 99%
18.02	Lafayette	5	Cold Spot 99%
19.01	Lafayette	7	Cold Spot 99%
19.02	Lafayette	1	Cold Spot 99%
19.03	Lafayette	9	Cold Spot 99%
19.04	Lafayette	1	Cold Spot 99%
19.05	Lafayette	2	Cold Spot 99%
2	Lafayette	3	Cold Spot 99%
20.01	Lafayette	9	Cold Spot 99%
20.02	Lafayette	4	Cold Spot 99%
202	St. Martin	13	Cold Spot 99%
203.01	St. Martin	1	Cold Spot 99%
203.02	St. Martin	2	Cold Spot 99%
204	St. Martin	2	Cold Spot 99%
205.01	St. Martin	1	Cold Spot 99%
205.02	St. Martin	4	Cold Spot 99%
206	St. Martin	3	Cold Spot 99%
208	St. Martin	5	Cold Spot 99%
209	St. Martin	1	Cold Spot 99%
21.01	Lafayette	4	Cold Spot 99%
21.02	Lafayette	2	Cold Spot 99%

(Table Cont'd)

Census Tract ID	Parish	Positive Count	Confidence Result
21.03	Lafayette	4	Cold Spot 99%
21.04	Lafayette	1	Cold Spot 99%
22	Lafayette	2	Cold Spot 99%
303	Avoyelles	1	Cold Spot 99%
303.01	Iberia	5	Cold Spot 99%
304	Iberia	2	Cold Spot 99%
305	Iberia	1	Cold Spot 99%
308	Iberia	2	Cold Spot 99%
309	Avoyelles	1	Cold Spot 99%
310	Iberia	2	Cold Spot 99%
311	Iberia	3	Cold Spot 99%
312	Iberia	2	Cold Spot 99%
313	Iberia	5	Cold Spot 99%
316	Iberia	3	Cold Spot 99%
6	Jefferson Davis	1	Cold Spot 99%
6.03	Lafayette	2	Cold Spot 99%
6.04	Lafayette	1	Cold Spot 99%
7	Lafayette	3	Cold Spot 99%
8	Lafayette	2	Cold Spot 99%
9	Lafayette	3	Cold Spot 99%
9501	Evangeline	3	Cold Spot 99%
9501	Vermilion	1	Cold Spot 99%
9503	Evangeline	1	Cold Spot 99%
9504	Evangeline	1	Cold Spot 99%
9505	Vermilion	1	Cold Spot 99%
9506	Evangeline	1	Cold Spot 99%
9506	Vermilion	2	Cold Spot 99%
9507	Evangeline	1	Cold Spot 99%
9508	Vermilion	1	Cold Spot 99%
9509.02	Vermilion	2	Cold Spot 99%
9510.01	Vermilion	1	Cold Spot 99%
9510.02	Vermilion	1	Cold Spot 99%
9601	Acadia	4	Cold Spot 99%
21.04	Lafayette	1	Cold Spot 99%
22	Lafayette	2	Cold Spot 99%
303	Avoyelles	1	Cold Spot 99%
303.01	Iberia	5	Cold Spot 99%
304	Iberia	2	Cold Spot 99%
305	Iberia	1	Cold Spot 99%
308	Iberia	2	Cold Spot 99%
309	Avoyelles	1	Cold Spot 99%
310	Iberia	2	Cold Spot 99%

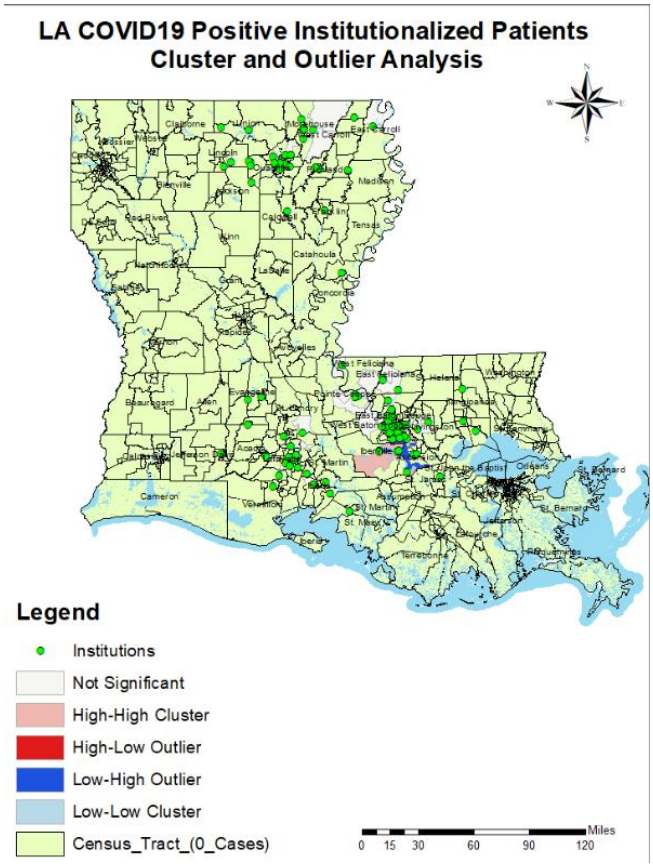
(Table Cont'd)

Census Tract ID	Parish	Positive Count	Confidence Result
311	Iberia	3	Cold Spot 99%
312	Iberia	2	Cold Spot 99%
313	Iberia	5	Cold Spot 99%
316	Iberia	3	Cold Spot 99%
6	Jefferson Davis	1	Cold Spot 99%
6.03	Lafayette	2	Cold Spot 99%
6.04	Lafayette	1	Cold Spot 99%
7	Lafayette	3	Cold Spot 99%
8	Lafayette	2	Cold Spot 99%
9	Lafayette	3	Cold Spot 99%
9501	Evangeline	3	Cold Spot 99%
9501	Vermilion	1	Cold Spot 99%
9503	Evangeline	1	Cold Spot 99%
9504	Evangeline	1	Cold Spot 99%
9505	Vermilion	1	Cold Spot 99%
9506	Evangeline	1	Cold Spot 99%
9506	Vermilion	2	Cold Spot 99%
9507	Evangeline	1	Cold Spot 99%
9508	Vermilion	1	Cold Spot 99%
9509.02	Vermilion	2	Cold Spot 99%
9510.01	Vermilion	1	Cold Spot 99%
9510.02	Vermilion	1	Cold Spot 99%
9601	Acadia	4	Cold Spot 99%
9602	Acadia	3	Cold Spot 99%
9603	Acadia	2	Cold Spot 99%
312	Iberia	2	Cold Spot 99%
313	Iberia	5	Cold Spot 99%
316	Iberia	3	Cold Spot 99%
6	Jefferson Davis	1	Cold Spot 99%
6.03	Lafayette	2	Cold Spot 99%
6.04	Lafayette	1	Cold Spot 99%
7	Lafayette	3	Cold Spot 99%
8	Lafayette	2	Cold Spot 99%
9	Lafayette	3	Cold Spot 99%
9501	Evangeline	3	Cold Spot 99%
9501	Vermilion	1	Cold Spot 99%
9503	Evangeline	1	Cold Spot 99%
9504	Evangeline	1	Cold Spot 99%
9505	Vermilion	1	Cold Spot 99%
9506	Evangeline	1	Cold Spot 99%
9506	Vermilion	2	Cold Spot 99%
9507	Evangeline	1	Cold Spot 99%

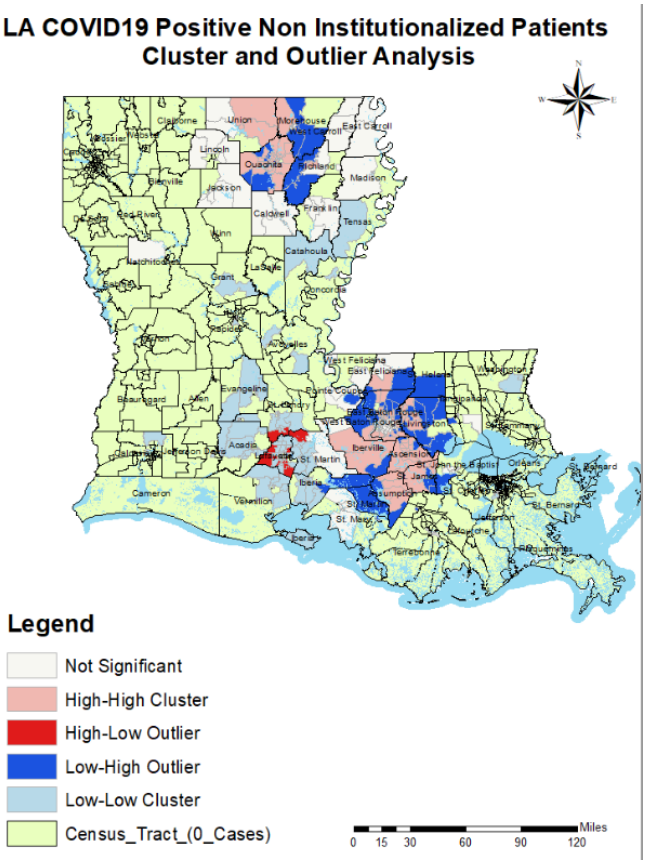
(Table Cont'd)

Census Tract ID	Parish	Positive Count	Confidence Result
9508	Vermilion	1	Cold Spot 99%
9509.02	Vermilion	2	Cold Spot 99%
9510.01	Vermilion	1	Cold Spot 99%
9510.02	Vermilion	1	Cold Spot 99%
9601	Acadia	4	Cold Spot 99%
9602	Acadia	3	Cold Spot 99%
9603	Acadia	2	Cold Spot 99%
9604	Acadia	2	Cold Spot 99%
9605	St. Landry	3	Cold Spot 99%
9606	Acadia	2	Cold Spot 99%
9607	Acadia	5	Cold Spot 99%
9608	Acadia	1	Cold Spot 99%
9608	St. Landry	1	Cold Spot 99%
9609	Acadia	3	Cold Spot 99%
9610	Acadia	1	Cold Spot 99%
9611	St. Landry	3	Cold Spot 99%
9612	St. Landry	1	Cold Spot 99%
9613	St. Landry	1	Cold Spot 99%
9614	St. Landry	1	Cold Spot 99%
9615	St. Landry	1	Cold Spot 99%
9617	St. Landry	1	Cold Spot 99%
9618	St. Landry	8	Cold Spot 99%
9619	St. Landry	3	Cold Spot 99%

Appendix AF. Cluster and Outlier Analysis of FMOLHS COVID Institutionalized and Non Institutionalized Positive Tested Patients



a) Institutionalized



b) Non-Institutionalized

Appendix AG. Map Summary tables of Cluster and Outlier Analysis

Map Summary table of Institutionalized Positive Patients

Results	Area (Kilometer Square)	# of Census Tracts	# of COVID-19 Positive	Total Population	# COVID-19/total pop x 1000
Cluster High-High	727.33	6	92	41493	2.22
Outlier Low-High	289.48	6	15	47661	0.31
Cluster Low-Low	15.99	1	4	8004	0.50
Not Significant	4282.36	58	443	277094	1.60

Map Summary Table of Non-Institutionalized Positive Patients

Results	Area (Kilometer Square)	# of Census Tracts	# of COVID-19 Positive	Total Population	# COVID-19/total pop x 1000
Cluster High-High	7073.04	92	1293	550087	2.35
Outlier High-Low	553.88	11	105	90607	1.16
Outlier Low-High	7685.54	64	210	301195	0.70
Cluster Low-Low	8738.46	97	207	474513	0.44
Not Significant	13494.79	100	388	408145	0.95

Appendix AH. Cluster and Outlier Analysis Percentage Averages of SVI and ACS Data

Results Confidence Spots/ SVI data % averages	High-High	High-Low	Low-High	Low-Low
Poverty	21.34	15.14	19.21	23.78
Unemployment	7.28	4.64	6.90	7.82
Per Capita Income	25156	30694	26408	23008
No High School Dip	14.44	13.00	16.70	19.98
Age 65	13.13	10.52	15.39	14.26
Age 17	25.30	26.57	24.06	25.18
Disability	14.19	11.42	15.46	15.59
Single Parent	13.37	11.89	11.29	12.36
Minority	51.44	31.56	36.43	42.91
Limited English Proficiency	0.93	1.95	0.93	1.65
Housing 10 or More	5.47	8.24	3.57	5.18
Mobile Homes	13.02	16.39	19.60	15.17
Crowded Housing	2.87	3.55	2.94	3.04
No Vehicle	8.22	4.56	7.32	10.58
Institutionalized Group Quarters	3.38	0.25	2.26	2.27
Socioeconomic Rank	0.47	0.33	0.44	0.58
Household Composition Rank	0.51	0.35	0.51	0.56
Minority/Language Rank	0.50	0.59	0.36	0.52
Housing/Transportation Rank	0.52	0.48	0.48	0.61
Overall Rank	0.50	0.39	0.44	0.60
Uninsured	10.81	10.43	10.07	12.81
Management, Business, Science & Art Occupations	30.80	36.01	29.16	27.40
Service Occupations	20.68	15.94	18.67	22.26
Sales & Office Occupations	23.44	22.23	23.36	22.01
Natural resources, construction, and maintenance occupations	10.11	14.36	13.22	13.19
Production, transportation, and material moving occupations	14.95	11.48	15.53	15.14

Appendix AI. Cluster and Outlier Analysis FMOLHS Positive Institutionalized Patients Counts By Census Tracts across LA

Census Tract ID	Parish	Positive Count	Result
11.02	East Baton Rouge	6	LH
12	Lafayette	4	LL
17	East Baton Rouge	20	HH
302.04	Ascension	1	LH
303	Ascension	5	LH
310	Ascension	15	HH
38.04	East Baton Rouge	1	LH
40.16	East Baton Rouge	1	LH
45.09	East Baton Rouge	21	HH
45.1	East Baton Rouge	15	HH
9529	Iberville	8	HH
9531.02	Iberville	1	LH
9532	Iberville	13	HH

Appendix AJ. Cluster and Outlier Analysis FMOLHS Positive Non-Institutionalized Patients Counts By Census Tracts across LA

Census Tract ID	Parish	Positive Count	Result
1	Ouachita	13	HH
101.01	Ouachita	9	HH
101.02	Ouachita	20	HH
102.01	Ouachita	21	HH
102.02	Ouachita	24	HH
103.01	Ouachita	10	HH
103.02	Ouachita	8	HH
104	Ouachita	10	HH
105.04	Ouachita	12	HH
106.03	Ouachita	14	HH
106.04	Ouachita	14	HH
107	Ouachita	7	HH
108	Ouachita	8	HH
109	Ouachita	22	HH
110	Ouachita	18	HH
111	Ouachita	11	HH
14	Ouachita	12	HH
17	Ouachita	14	HH
2	East Baton Rouge	12	HH
2	Ouachita	14	HH
204.01	West Baton Rouge	16	HH
204.02	West Baton Rouge	14	HH
301.01	Ascension	9	HH
301.02	Ascension	14	HH
302.04	Ascension	21	HH
302.05	Ascension	24	HH
302.06	Ascension	22	HH
303	Ascension	62	HH
304.01	Ascension	22	HH
304.02	Ascension	57	HH
305	Ascension	8	HH
306	Ascension	25	HH
309	Ascension	17	HH
31.01	East Baton Rouge	8	HH
310	Ascension	40	HH
32.01	East Baton Rouge	7	HH
33	East Baton Rouge	16	HH
34	East Baton Rouge	18	HH
35.04	East Baton Rouge	16	HH

(Table cont'd)

Census Tract ID	Parish	Positive Count	Result
35.05	East Baton Rouge	13	HH
35.07	East Baton Rouge	14	HH
4	East Baton Rouge	11	HH
4.01	Ouachita	7	HH
4.02	Ouachita	7	HH
40.06	East Baton Rouge	12	HH
40.09	East Baton Rouge	8	HH
40.11	East Baton Rouge	9	HH
40.14	East Baton Rouge	7	HH
40.15	East Baton Rouge	14	HH
40.16	East Baton Rouge	9	HH
401	St. James	10	HH
403.01	Livingston	8	HH
403.03	Livingston	7	HH
404	St. James	14	HH
404.02	Livingston	7	HH
405	Livingston	7	HH
405	St. James	17	HH
406	Livingston	7	HH
408.04	Livingston	15	HH
42.01	East Baton Rouge	12	HH
42.05	East Baton Rouge	11	HH
44.01	East Baton Rouge	7	HH
45.03	East Baton Rouge	8	HH
45.04	East Baton Rouge	8	HH
45.05	East Baton Rouge	9	HH
45.1	East Baton Rouge	9	HH
46.02	East Baton Rouge	8	HH
46.04	East Baton Rouge	13	HH
5	Ouachita	8	HH
501	Assumption	29	HH
503	Assumption	38	HH
504	Assumption	24	HH
505	Assumption	11	HH
52.04	Ouachita	17	HH
53.01	Ouachita	12	HH
54	Ouachita	8	HH
55	Ouachita	9	HH
58	Ouachita	7	HH
6	Ouachita	18	HH
6.01	East Baton Rouge	7	HH
9	Ouachita	8	HH
9503	Morehouse	7	HH

(Table cont'd)

Census Tract ID	Parish	Positive Count	Result
9515.02	East Feliciana	9	HH
9527	Iberville	12	HH
9529	Iberville	22	HH
9531.01	Iberville	10	HH
9531.02	Iberville	7	HH
9532	Iberville	15	HH
9601	Union	12	HH
9604	Union	7	HH
9605	Union	8	HH
9705	Richland	11	HH
12	Lafayette	9	HL
14.01	Lafayette	9	HL
14.04	Lafayette	10	HL
14.05	Lafayette	7	HL
14.06	Lafayette	10	HL
14.1	Lafayette	14	HL
19.01	Lafayette	7	HL
19.03	Lafayette	9	HL
20.01	Lafayette	9	HL
202	St. Martin	13	HL
9618	St. Landry	8	HL
1	East Baton Rouge	4	LH
105.02	Ouachita	1	LH
105.03	Ouachita	3	LH
11	Ouachita	4	LH
15	Ouachita	6	LH
202	West Baton Rouge	2	LH
203	West Baton Rouge	4	LH
205	Lafourche	1	LH
210	St. Martin	1	LH
3	East Baton Rouge	6	LH
30	East Baton Rouge	3	LH
301	Iberia	4	LH
301.03	Ascension	4	LH
302.03	Ascension	5	LH
32.02	East Baton Rouge	5	LH
35.01	East Baton Rouge	5	LH
36.03	East Baton Rouge	4	LH
39.06	East Baton Rouge	3	LH
40.1	East Baton Rouge	2	LH
401	Livingston	1	LH
402	St. James	4	LH

(Table cont'd)

Census Tract ID	Parish	Positive Count	Result
402	St. Mary	1	LH
402.01	Livingston	2	LH
402.02	Livingston	2	LH
403	St. James	3	LH
403.04	Livingston	5	LH
404.01	Livingston	5	LH
406	St. James	6	LH
407	Livingston	4	LH
407	St. James	2	LH
408.02	Livingston	4	LH
408.05	Livingston	3	LH
408.06	Livingston	5	LH
409.02	Livingston	4	LH
42.04	East Baton Rouge	2	LH
43.01	East Baton Rouge	5	LH
43.02	East Baton Rouge	2	LH
44.02	East Baton Rouge	4	LH
44.03	East Baton Rouge	3	LH
45.08	East Baton Rouge	6	LH
45.09	East Baton Rouge	6	LH
46.03	East Baton Rouge	4	LH
47	East Baton Rouge	5	LH
502	Assumption	3	LH
506	Assumption	5	LH
51	Ouachita	4	LH
52.01	Ouachita	1	LH
52.03	Ouachita	5	LH
53.02	Ouachita	6	LH
59	Ouachita	1	LH
7	Ouachita	1	LH
7.01	East Baton Rouge	2	LH
711	St. John the Baptist	1	LH
9502	Morehouse	5	LH
9506	Morehouse	4	LH
9507	Morehouse	1	LH
9512	St. Helena	1	LH
9515.01	East Feliciana	3	LH
9516	East Feliciana	3	LH
9522	Pointe Coupee	2	LH
9530	Iberville	3	LH
9541.02	Tangipahoa	1	LH
9704	Richland	2	LH

(Table cont'd)

Census Tract ID	Parish	Positive Count	Result
9706	Richland	1	LH
1	Catahoula	1	LL
1	Lafayette	3	LL
10.01	Lafayette	5	LL
10.02	Lafayette	2	LL
10.03	Lafayette	3	LL
11	Lafayette	6	LL
120	Rapides	1	LL
125	Rapides	1	LL
13	Lafayette	2	LL
14.02	Lafayette	5	LL
14.03	Lafayette	6	LL
14.07	Lafayette	2	LL
14.09	Lafayette	5	LL
17	Lafayette	3	LL
17.25	Orleans	1	LL
18.01	Lafayette	1	LL
18.02	Lafayette	5	LL
19.02	Lafayette	1	LL
19.04	Lafayette	1	LL
19.05	Lafayette	2	LL
2	Concordia	2	LL
2	Lafayette	3	LL
2	Tensas	2	LL
20.02	Lafayette	4	LL
203.01	St. Martin	1	LL
203.02	St. Martin	2	LL
204	St. Martin	2	LL
204.01	Grant	1	LL
205.01	St. Martin	1	LL
205.02	St. Martin	4	LL
206	St. Martin	3	LL
208	St. Martin	5	LL
209	St. Martin	1	LL
21.01	Lafayette	4	LL
21.02	Lafayette	2	LL
21.03	Lafayette	4	LL
21.04	Lafayette	1	LL
218.04	Jefferson	1	LL
22	Lafayette	2	LL
22.01	Calcasieu	1	LL
25.03	Orleans	1	LL

(Table cont'd)

Census Tract ID	Parish	Positive Count	Result
26	Calcasieu	1	LL
269	Jefferson	1	LL
277.01	Jefferson	1	LL
303	Avoyelles	1	LL
303.01	Iberia	5	LL
304	Iberia	2	LL
305	Iberia	1	LL
306	Iberia	2	LL
308	Iberia	2	LL
309	Avoyelles	1	LL
310	Iberia	2	LL
311	Iberia	3	LL
312	Iberia	2	LL
313	Iberia	5	LL
316	Iberia	3	LL
410.03	St. Tammany	1	LL
6	Jefferson Davis	1	LL
6.03	Lafayette	2	LL
6.04	Lafayette	1	LL
7	Lafayette	3	LL
8	Lafayette	2	LL
9	Lafayette	3	LL
9501	Evangeline	3	LL
9501	Vermilion	1	LL
9503	Evangeline	1	LL
9504	Evangeline	1	LL
9505	Vermilion	1	LL
9506	Evangeline	1	LL
9506	Vermilion	2	LL
9506	Washington	1	LL
9507	Evangeline	1	LL
9507	Washington	1	LL
9508	Vermilion	1	LL
9509.02	Vermilion	2	LL
9510.01	Vermilion	1	LL
9510.02	Vermilion	1	LL
9601	Acadia	4	LL
9602	Acadia	3	LL
9603	Acadia	2	LL
9604	Acadia	2	LL
9604	Madison	1	LL
9605	Madison	2	LL

(Table cont'd)

Census Tract ID	Parish	Positive Count	Result
9605	St. Landry	3	LL
9606	Acadia	2	LL
9607	Acadia	5	LL
9608	Acadia	1	LL
9608	St. Landry	1	LL
9609	Acadia	3	LL
9610	Acadia	1	LL
9611	St. Landry	3	LL
9612	St. Landry	1	LL
9613	St. Landry	1	LL
9614	St. Landry	1	LL
9615	St. Landry	1	LL
9617	St. Landry	1	LL
9619	St. Landry	3	LL

References

1. About Us. Franciscan Missionaries of Our Lady Health System; Available from: https://fmolhs.org/about-us/. Accessed 28 December 2019

2. About Us. Our Lady of the Lake Regional Medical Center; Available from: https://ololrmc.com/about-us/. Accessed 28 December 2019

3. Services. Our Lady of the Lake Regional Medical Center; Available from: https://ololrmc.com/services/. Accessed 28 December 2019

4. LSU Healthcare Network | New Orleans Specialty & Primary Care Center. LSU Health New Orleans 2019; Available from: http://www.lsuhn.com/.

5. Hayes, T.O.N. PRIMER: Examining Trends in Emergency Department Utilization and Costs. AAF 2018; Available from: https://www.americanactionforum.org/research/primer-examining-trends-in-emergency-department-utilization-and-costs/.

6. Weinick, R.M., R.M. Burns, and A. Mehrotra, *Many Emergency Department Visits Could Be Managed At Urgent Care Centers And Retail Clinics*. 2010. p. 1630-1636.

7. Moore, B.J., C. Stocks, and P.L. Ownens, *Trends in Emergency Department Visits*, 2006–2014. Agency for Healthcare Research and Quality, 2017.

8. Lee, D.C., et al., *The Impact of Hospital Closures and Hospital and Population Characteristics on Increasing Emergency Department Volume*: A Geographic Analysis. 2015. p. 459-466.

9. Erik, S.A., et al., *Addressing Social Determinants of Health from the Emergency Department through Social Emergency Medicine*. Western Journal of Emergency Medicine, 2016(4): p. 487.

10. Public Health Response to the Initiation and Spread of Pandemic COVID-19 in the United States, February 24–April 21, 2020. Centers for Disease Control and Prevention May 07, 2020 May 07, 2020; Available from: https://www.cdc.gov/mmwr/volumes/69/wr/mm6918e2.htm.

11. Cases in the U.S. Centers for Disease Control and Prevention July 11, 2020; Available from: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/cases-in-us.html.

12. Zhang Y. The Epidemiological Characteristics of an Outbreak of 2019 Novel Coronavirus Disease (COVID-19) - China, 2020. China CDC Weekly. 2020;2(8):113 - 122.

13. Team CC-R. Severe Outcomes Among Patients with Coronavirus Disease 2019 (COVID-19) - United States, February 12-March 16, 2020. MMWR Morb Mortal Wkly Rep. 2020;69(12):343-346.

14. “Do I Need to Take Extra Precautions Against COVID-19.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/index.html.

15. Team CC-R. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 - United States, February 12-March 28, 2020. MMWR Morb Mortal Wkly Rep. 2020;69(13):382-386.

16. Arentz M, Yim E, Klaff L, et al. Characteristics and Outcomes of 21 Critically Ill Patients With COVID-19 in Washington State. JAMA. 2020.

17. Buckner FS, McCulloch DJ, Atluri V, et al. Clinical Features and Outcomes of 105 Hospitalized patients with COVID-19 in Seattle, Washington. Clin Infect Dis. 2020.

18. Richardson S, Hirsch JS, Narasimhan M, et al. Presenting Characteristics, Comorbidities, and Outcomes Among 5700 Patients Hospitalized With COVID-19 in the New York City Area. JAMA. 2020.

19. Abrams, Elissa M, and Stanley J Szefler. “COVID-19 and the Impact of Social Determinants of Health.” The Lancet. Respiratory Medicine, Elsevier Ltd., July 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7234789/.

20. Federico MJ, McFarlane AE, 2nd, Szefler SJ, Abrams EM. The impact of social determinants of health on children with asthma. J Allergy Clin Immunol Pract. 2020 doi: 10.1016/j.jaip.2020.03.028. published online April 12.

21. Cases & Deaths by County. Centers for Disease Control and Prevention June 26, 2020; Available from: https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/county-map.html?state=LA.

22. Social determinants of health. World Health Organization 2019; Available from: https://www.who.int/social_determinants/sdh_definition/en/.

23. Padarthy, S., K. Knudson, and S. Vattikuti, *The Social Determinantsof Health: Applying AI & Machine Learning to Achieve Whole Person Care* 2019, Cognizant.

24. Best States Rankings. U.S. News & World Report 2019; Available from: https://www.usnews.com/news/best-states/rankings.

25. Brunt, D.V., *Community Health Records: Establishing a Systematic Approach to Improving Social and Physical Determinants of Health*. Am J Public Health, 2017. 107: p. 407-412.

26. Social Determinants of Health. Social Determinants of Health | Healthy People 2020; Available from: https://www.healthypeople.gov/2020/topics-objectives/topic/social-determinants-of-health.

27. Trzeciak, S., et al., *Emergency department overcrowding in the United States: an emerging threat to patient safety and public health*. Emergency Medicine Journal, 2003: p. 402-405.

28. Uscher-Pines, L., et al., *Emergency department visits for nonurgent conditions: systematic literature review*. U.S. National Library of Medicine.

29. Claire, M., et al., *Emergency department crowding: A systematic review of causes, consequences and solutions*. PLoS ONE, 2018(8): p. e0203316.

30. Hines, A.L., et al., *Conditions With the Largest Number of Adult Hospital Readmissions by Payer*, 2011: Statistical Brief #172. Healthcare Cost and Utilization Project (HCUP) Statistical Briefs, 2006.

31. Shuler, M. Baton Rouge General to close its Mid City emergency room. The Advocate 2015; Available from: https://www.theadvocate.com/baton_rouge/news/politics/legislature/article_dc9916d9-5ad5-5735-b610-4c302102a497.html.

32. Louisiana Coronavirus COVID-19: Department of Health: State of Louisiana. Louisiana Coronavirus COVID-19 | Department of Health | State of Louisiana July 14, 2020; Available from: http://ldh.la.gov/coronavirus/.

33. COVID-19 in Racial and Ethnic Minority Groups. Centers for Disease Control and Prevention June 25, 2020; Available from: https://www.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/racial-ethnic-minorities.html.

34. Provisional COVID-19 Death Counts by Sex, Age, and State. Centers for Disease Control and Prevention July 14, 2020; Available from: https://data.cdc.gov/NCHS/Provisional-COVID-19-Death-Counts-by-Sex-Age-and-S/9bhg-hcku.

35. Jianmin, W., et al., *A practical method for predicting frequent use of emergency department care using routinely available electronic registration data*. BMC Emergency Medicine, 2016. 16: p. 1-9.

36. Pereira, M., et al., *Predicting Future Frequent Users of Emergency Departments in California State. Proceedings of the 5th ACM Conference on Bioinformatics*, Computational Biology & Health Informatics, 2016: p. 603.

37. Hatef, E., J.P. Weiner, and H. Kharrazi, *A public health perspective on using electronic health records to address social determinants of health*: *The potential for a national system of local community health records in the United States*. International Journal of Medical Informatics, 2019. 124: p. 86-89.

38. QIAO, Z., et al., *Using Machine Learning Approaches for Emergency Room Visit Prediction ased on Electronic Health Record Data*. European Federation for Medical Informatics (EFMI) and IOS Press, 2018: p. 111-115.

39. Chengliang, Y., et al., *Machine learning approaches for predicting high cost high need patient expenditures in health care. BioMedical Engineering OnLine*, 2018(S1): p. 1.

40. Graham, B., et al., *Using Data Mining to Predict Hospital Admissions From the Emergency Department*. IEEE Access, Access, IEEE, 2018: p. 10458.

41. Vest, J.R. and O. Ben-Assuli, *Prediction of emergency department revisits using area-level social determinants of health measures and health information exchange information*. International Journal of Medical Informatics, 2019. 129: p. 205-210.

42. Agarwal, R. The 5 Classification Evaluation metrics every Data Scientist must know. Towards Data Science 2019; Available from: https://towardsdatascience.com/the-5-classification-evaluation-metrics-you-must-know-aa97784ff226.

43. ICD - ICD-9-CM - International Classification of Diseases, Ninth Revision, Clinical Modification. Centers for Disease Control and Prevention Nov 2015; Available from: https://www.cdc.gov/nchs/icd/icd9cm.htm#ftp.

44. Grinspan, Z.M., et al., *Predicting frequent emergency department use among children with epilepsy: A retrospective cohort study using electronic health data from 2 centers. Epilepsia*, 2018. 59(1): p. 155-169.

45. Tang, F., et al., *Predictive modeling in urgent care: a comparative study of machine learning approaches*. JAMIA Open, 2018. 1(1): p. 87-98.

46. Muhammet, G. and G. Ali Fuat, *FORECASTING PATIENT LENGTH OF STAY IN AN EMERGENCY DEPARTMENT BY ARTIFICIAL NEURAL NETWORKS*. Havacılık ve Uzay Teknolojileri Dergisi, 2015(2): p. 43.

47. Benjamin, S., T. Shripad, and R. David, *Machine learning approaches to the social determinants of health in the health and retirement study*. SSM: Population Health, 2018(95-99): p. 95.

48. Sukel, K., *Nonclinical Factors Impact Patient Outcomes: Social determinants of health play growing role in hospital, payer strategies*. Managed Healthcare Executive, 2018. 28(12): p. 34-37.

49. Pham, J.C., J.D. Bayram, and D.K. Moss. Characteristics of Frequent Users of Three Hospital Emergency Departments. AHRQ 2017; Available from: https://www.ahrq.gov/professionals/systems/hospital/edenvironmentalscan/researchbrief.html.

50. Chambers, C., et al., *High Utilizers of Emergency Health Services in a Population-Based Cohort of Homeless Adults. American Journal of Public Health*, 2013. 103(S2): p. S302-S310.

51. Hatef, E., et al., *The Impact of Social Determinants of Health on Hospitalization in the Veterans Health Administration*. American Journal of Preventive Medicine, 2019. 56(6): p. 811-818.

52. Hwang, S.W., et al., *A Comprehensive Assessment of Health Care Utilization Among Homeless Adults Under a System of Universal Health Insurance*. American Journal of Public Health, 2013. 103(S2): p. S294-S301.

53. Doran, K.M., et al., *Homelessness and other social determinants of health among emergency department patients*. Journal of Social Distress & the Homeless, 2016. 25(2): p. 71-77.

54. Key Statistics & Graphics. USDA ERS - Key Statistics & Graphics 2019; Available from: https://www.ers.usda.gov/topics/food-nutrition-assistance/food-security-in-the-us/key-statistics-graphics.aspx.

55. State of Homelessness. National Alliance to End Homelessness; Available from: https://endhomelessness.org/homelessness-in-america/homelessness-statistics/state-of-homelessness-report/.

56. Report on the Economic Well-Being of U.S. Households in 2018 - May 2019. BOARD OF GOVERNORS of the FEDERAL RESERVE SYSTEM 2019; Available from: https://www.federalreserve.gov/publications/2019-economic-well-being-of-us-households-in-2018-preface.htm.

57. Heath, S. SDOH Lead to Poor Patient Outcomes, High Healthcare Utilization. Xtelligent Healthcare Media May 2019; Available from: https://patientengagementhit.com/news/sdoh-lead-to-poor-patient-outcomes-high-healthcare-utilization.

58. Axelson, D.J., M.J. Stull, and W.C. Coates, *Social Determinants of Health: A Missing Link in Emergency Medicine Training*. AEM Education And Training, 2017. 2(1): p. 66-68.

59. Miner, J.R., et al., *Hunger and Food Insecurity among Patients in an Urban Emergency Department. The Western Journal Of Emergency Medicine*, 2013. 14(3): p. 253-262.

60. Griffey, R.T., et al., *Is Low Health Literacy Associated With Increased Emergency Department Utilization and Recidivism?* (vol 21, pg 1109, 2014). 2015. p. 497-497.

61. Runyan, C.N., *Assessing social determinants of health in primary care: Liability or opportunity?* Families, Systems, & Health, 2018. 36(4): p. 550-552.

62. Greg, S., *Boston Medical Center implements SDOH screening tool in EHR*. 2019.

63. Fenton. Health Care's Blind Side. 2011; Available from: https://www.rwjf.org/content/rwjf/en/library/research/2011/12/health-care-s-blind-side.html.

64. Price-Haywood, E.G., et al., *Hospitalization and Mortality among Black Patients and White Patients with Covid-19.* New England Journal of Medicine, 2020. 382(26): p. 2534-2543.

65. Kim, S.J. and W. Bostwick, *Social Vulnerability and Racial Inequality in COVID-19 Deaths in Chicago*. Health education & behavior : the official publication of the Society for Public Health Education, 2020. 47(4): p. 509-513.

66. Khalatbari-Soltani, S., et al., *Importance of collecting data on socioeconomic determinants from the early stage of the COVID-19 outbreak onwards*. Journal of epidemiology and community health, 2020. 74(8): p. 620-623.

67. CDC's Social Vulnerability Index (SVI). Centers for Disease Control and Prevention; Available from: https://svi.cdc.gov/data-and-tools-download.html.

68. Rural-Urban Commuting Area Codes. USDA ERS - Rural-Urban Commuting Area Codes; Available from: https://www.ers.usda.gov/data-products/rural-urban-commuting-area-codes/.

69. data.census.gov. data.census.gov; Available from: https://data.census.gov/cedsci/table?q=DP03.

70. Goldberg, D.W. Texas A&M University Geoservices. Available from: http://geoservices.tamu.edu.

71. How Spatial Autocorrelation (Global Moran's I) works. How Spatial Autocorrelation (Global Moran's I) Works-ArcGIS Pro | ArcGIS Desktop; Available from: https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-spatial-autocorrelation-moran-s-i-spatial-st.htm.

72. Dahiru, T., *P - value, a true test of statistical significance? A cautionary note*. Annals Of Ibadan Postgraduate Medicine, 2008. 6(1): p. 21-26.

73. How Hot Spot Analysis (Getis-Ord Gi*) works. How Hot Spot Analysis (Getis-Ord Gi*) works-ArcGIS Pro | Documentation; Available from: https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/h-how-hot-spot-analysis-getis-ord-gi-spatial-stati.htm.

74. *Cluster and Outlier Analysis (Anselin Local Moran's I) (Spatial Statistics)*. Cluster and Outlier Analysis (Anselin Local Moran's I) (Spatial Statistics)-ArcGIS Pro | Documentation; Available from: https://pro.arcgis.com/en/pro-app/tool-reference/spatial-statistics/cluster-and-outlier-analysis-anselin-local-moran-s.htm.

75. *Learn to Use the Eta Coefficient Test in SPSS With Data From the NIOSH Quality of Worklife Survey (2014).* 2019: London.

76. van Kuijk, S.M.J., et al., *Preparing Data for Predictive Modelling*. Fundamentals of Clinical Data Science, 2019: p. 75-84.

77. Dreyfus, G., *Neural networks. [electronic resource] : methodology and applications*. 2005: Springer.

78. *File:Neural Network Example.svg*. en.wikipedia.org/wiki/File:Neural_network_example.svg.

79. *Sklearn.neural_network.MLPClassifier*. scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html.

80. Suthaharan, S., *Machine learning models and algorithms for big data classification : thinking with examples for effective learning.* Integrated series in information systems: Volume 36. 2016: Springer Science+Business Media.

81. Nikulski, Julia. *The Ultimate Guide to AdaBoost, Random Forests and XGBoost*. 7 Oct. 2020, towardsdatascience.com/the-ultimate-guide-to-adaboost-random-forests-and-xgboost-7f9327061c4f.

82. Ramzai, Juhi. *Simple Guide for Ensemble Learning Methods*. 6 Mar. 2019, towardsdatascience.com/simple-guide-for-ensemble-learning-methods-d87cc68705a2#:~:text=2.-,Boosting,and%20builds%20strong%20predictive%20models.

83. Sarkar, Priyankur, and KnowledgeHut. *Bagging and Random Forest in Machine Learning: How Do They Work?* 14 Oct. 2019, www.knowledgehut.com/blog/data-science/bagging-and-random-forest-in-machine-learning.

84. *Sklearn.ensemble.RandomForestClassifier*. scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html.

85. *(PDF) Experimenting XGBoost Algorithm for Prediction and* www.researchgate.net/publication/318132203_Experimenting_XGBoost_Algorithm_for_Prediction_and_Classification_of_Different_Datasets.

86. XGBoost Parameters. xgboost.readthedocs.io/en/latest/parameter.html.

87. Data.Medicaid.gov. Medicaid.gov; Available from: https://data.medicaid.gov/Enrollment/State-Medicaid-and-CHIP-Applications-Eligibility-D/n5ce-jxme.

88. data.census.gov. Available from: https://data.census.gov/cedsci/table?q=race%20percentages%20in%20la&tid=ACSDP5Y2018.DP05&hidePreview=true.

Vita

RamyaKrishna Tummala received a Bachelor of Science degree in Mechanical Engineering from Velagapudi Ramakrishna Siddhartha Engineering College affiliated to Nagarjuna University, India in May 2013. She is pursuing her Master's degree in Engineering Science in Information Technology Engineering. As part of her thesis study, she worked on predictive modelling of Franciscan Missionaries of Our Lady Health System (FMOLHS) emergency room utilization using machine learning algorithms. Additionally, she performed COVID data analysis with the FMOLHS patient data that can help hospitals develop a strategic response throughout the different stages of pandemic under the direction of Dr. Gerald Knapp at LSU. Her research interests include data analytics and geospatial analysis. She is expected to graduate in May 2021.